



## Estimating functions for jump–diffusions

Jakobsen, Nina Munkholt; Sørensen, Michael

*Published in:*  
Stochastic Processes and Their Applications

*DOI:*  
[10.1016/j.spa.2018.09.006](https://doi.org/10.1016/j.spa.2018.09.006)

*Publication date:*  
2019

*Document version*  
Peer reviewed version

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Jakobsen, N. M., & Sørensen, M. (2019). Estimating functions for jump–diffusions. *Stochastic Processes and Their Applications*, 129, 3282–3318. <https://doi.org/10.1016/j.spa.2018.09.006>

# Estimating functions for jump-diffusions

Nina Munkholt Jakobsen <sup>\*1,2</sup> and Michael Sørensen <sup>†2</sup>

<sup>1</sup>Department of Applied Mathematics and Computer Science, Technical University of Denmark,  
Richard Petersens Plads, Building 324, DK-2800 Kgs. Lyngby, Denmark

<sup>2</sup>Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5,  
DK-2100 Copenhagen Ø, Denmark

November 10, 2018

## Abstract

Asymptotic theory for approximate martingale estimating functions is generalised to diffusions with finite-activity jumps, when the sampling frequency and terminal sampling time go to infinity. Rate optimality and efficiency are of particular concern. Under mild assumptions, it is shown that estimators of drift, diffusion, and jump parameters are consistent and asymptotically normal, as well as rate-optimal for the drift and jump parameters. Additional conditions are derived, which ensure rate-optimality for the diffusion parameter as well as efficiency for all parameters. The findings indicate a potentially fruitful direction for the further development of estimation for jump-diffusions.

**Keywords:** Approximate martingale estimating function, diffusion with jumps, discrete-time sampling, efficiency, optimal rate, stochastic differential equation.

**Running title:** Estimation for jump-diffusions.

---

<sup>\*</sup>nmja@dtu.dk

<sup>†</sup>michael@math.ku.dk

# 1 Introduction

When modelling phenomena in continuous time, diffusions with jumps are a natural generalisation or improvement of continuous diffusion processes driven by Wiener noise, or of pure-jump processes. Jump-diffusion models find application in, among other fields, biology [15], neuroscience [7, 12, 23, 47, 49], finance [6, 33, 46], and engineering [19]. Statistical inference for diffusions with jumps raises a broad spectrum of intriguing challenges. The models have continuous-time dynamics, but sampling in continuous time is not feasible. Furthermore, a closed-form expression for the likelihood function based on discrete-time observations is not available, rendering maximum likelihood estimation impracticable.

A number of estimation approaches based on discrete-time observations have previously been proposed in the literature. A non-exhaustive list of references includes the following. In the context of parametric estimation, pseudo-likelihood methods involving, primarily, Gaussian-inspired approximations of the log-likelihood function have been considered [36, 44, 45, 48, 52, 55], as well as closed-form expansion of the transition densities [8, 35, 63], and approximations to maximum likelihood estimators obtained from the continuous-time likelihood function [13, 37]. A quadratic variation-inspired estimation method was proposed in a semiparametric setting [39], and several non-parametric procedures [1, 40, 41, 50, 51, 53, 54, 62, 64] as well as a selection of simulation-based methods [11, 16, 17, 59] have been considered. Finally, parametric estimation for diffusions with jumps based on observations made in continuous time has been investigated too [56].

The present paper focuses on parametric estimation for an ergodic stochastic process  $\mathbf{X}^\theta = (X_t^\theta)_{t \geq 0}$  with finite-activity jumps, using discrete-time observations. The process is assumed to take values in the open interval  $\mathcal{X} \subseteq \mathbb{R}$ , and solve a stochastic differential equation (SDE) of the form

$$dX_t^\theta = a(X_t^\theta, \theta) dt + b(X_t^\theta, \theta) dW_t + \int_{\mathbb{R}} c(X_{t-}^\theta, z, \theta) N^\theta(dt, dz), \quad (1.1)$$

for  $\theta$  in an open parameter set  $\Theta \subseteq \mathbb{R}^d$ ,  $d \geq 1$ . The drift, diffusion, and jump coefficients, denoted  $a$ ,  $b$ , and  $c$ , respectively, are specified deterministic functions. As usual,  $\mathbf{X}_-^\theta = (X_{t-}^\theta)_{t \geq 0}$  is defined as the process of left limits of  $\mathbf{X}^\theta$ . The standard Wiener process  $(W_t)_{t \geq 0}$  is assumed to be independent of  $N^\theta(dt, dz)$ , a time-homogeneous, finite-activity Poisson random measure on  $[0, \infty) \times \mathbb{R}$ . Supposing that the stochastic process is observed at times  $t_i^n = i\Delta_n$ ,  $i = 0, 1, \dots, n$ ,  $\Delta_n > 0$ , we consider a high-frequency asymptotic scenario with an infinite time horizon:  $\Delta_n \rightarrow 0$  and  $n\Delta_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Assuming also the existence of a true, unknown parameter  $\theta_0 \in \Theta$ , we put  $\mathbf{X} = \mathbf{X}^{\theta_0}$  in the following, and introduce the notation  $X_{n,i}^\theta = X_{t_i^n}^\theta$  and  $X_i^n = X_{t_i^n}$ .

In [48, 55], two very similar contrast functions were proposed for estimation in models of the type (1.1) with finite-activity jumps. Both papers assume a parameter

$\theta$  separated into two components, one present in only the drift and jump terms of the SDE, the other only entering into the diffusion term. The corresponding estimators are rate optimal and efficient. An essential aspect of these contrast functions is the technique for deciding whether or not a jump is likely to have occurred in an observation interval  $[t_{i-1}^n, t_i^n]$ . This method has become standard in financial econometrics [34, 38]. Models similar to (1.1), but allowing also infinite-activity jumps, were treated in [44, 45]. In these models, the parameter separates into a drift component entering into only the drift term of the SDE, and a noise component figuring in both the diffusion and jump terms. For estimation, specific choices of Gaussian quasi-likelihood functions were used, which are known to work well for diffusions without jumps. In the presence of jumps, under an asymptotic scenario very similar to the one in the present paper, these Gaussian quasi-likelihood estimators were found to be neither rate optimal for the noise parameter nor efficient for any part of the parameter.

In this paper, we study approximate martingale estimating functions which may be written on the form

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^n g(\Delta_n, X_i^n, X_{i-1}^n, \theta). \quad (1.2)$$

The estimating function is defined by a deterministic function  $g(t, y, x, \theta)$ , which satisfies an approximate martingale property. This entails that for all  $\theta \in \Theta$ , the conditional expectation  $\mathbb{E}(g(\Delta_n, X_{n,i}^\theta, X_{n,i-1}^\theta, \theta) \mid X_{n,i-1}^\theta)$  is of order  $\Delta_n^{\kappa_0}$  for some constant  $\kappa_0 \geq 2$ . Estimators are obtained as solutions to the equation  $G_n(\theta) = 0$ , and referred to as  $G_n$ -estimators. For example, the Gaussian quasi-likelihood estimators considered in [44, 45] can, under regularity assumptions, be formulated in terms of approximate martingale estimating functions.

Approximate martingale estimating functions for continuous diffusions have already been quite thoroughly studied, see e.g. [4, 20, 21, 25, 30, 58, 61]. In particular, the existing theory includes high-frequency asymptotics for the estimators under an infinite time horizon [58] as well as infill asymptotics [25]. Both of these papers present simple conditions on the estimating functions which ensure rate optimality and efficiency. It should also be noted that a large part of the estimators proposed in the literature for continuous diffusions can be treated within the framework of approximate martingale estimating functions [58].

Compared to continuous diffusions, statistical theory for diffusions with jumps is still establishing itself, and many open questions are yet to be solved. Approximate martingale estimating functions constitute a tractable, rather comprehensive framework for the study of estimation for continuous diffusions. One would, therefore, expect research on approximate martingale estimating functions to provide fruitful insight into estimation theory for jump-diffusions. In this paper, we explore high-frequency asymptotics for general approximate martingale estimating functions, an area which has not previously been studied in the context of jump-diffusions.

First we establish, under mild assumptions, a general theorem on the existence and uniqueness of consistent, asymptotically normal  $G_n$ -estimators of the parameter  $\theta$  in the SDE model (1.1). We also provide a consistent estimator of the asymptotic variance. Next, we investigate the question of rate optimality and efficiency of the estimators of  $\theta$ . Usually, the optimal rate of convergence and efficient asymptotic variance would be identified using results from the theory of local asymptotic normality. However, local asymptotic normality and, for infill asymptotics, local asymptotic mixed normality are ongoing areas of research for stochastic processes with jumps [3, 5, 27, 28, 31, 32]. No results for general jump-diffusions have been established so far. Nonetheless, the optimal rates of convergence and the Fisher information matrices in the current setup are pretty clear. It can rather safely be conjectured that the optimal rate of convergence is  $\sqrt{n\Delta_n}$  for drift and jump components of the parameter and  $\sqrt{n}$  for diffusion components, and that the efficient asymptotic variance is as proposed in Section 4. These conjectures are motivated not only by local asymptotic normality results which cover particular submodels of (1.1) [3, 27, 31, 32], but also by other asymptotic results [14, 55, 56].

Considering two separate cases, we give conditions on  $g$  under which the corresponding approximate martingale estimating function  $G_n(\theta)$  yields rate optimal and efficient estimators. First, we assume the model (1.1) with no unknown parameter in the diffusion coefficient, so that there is only a drift-jump parameter to be estimated. Next, we present the case where (1.1) has a two-dimensional drift-jump parameter and a one-dimensional diffusion parameter. The restriction on the dimension of the parameter is due to the observation that when the diffusion coefficient depends on an unknown parameter, the complexity of the conditions obtained for rate optimality and efficiency increases substantially with the dimension of the parameter.

For our jump-diffusion models, in addition to the simple rate-optimality and efficiency conditions found by [58] in the context of continuous diffusions, several new conditions appear. An intuition for these results can be obtained by the following considerations. In the limit  $\Delta_n \rightarrow 0$  (asymptotically), a full sample path of  $\mathbf{X}$  is observed. In this hypothetical situation, all jump times may be identified as times  $t$  for which  $X_t \neq X_{t-}$ , with jump sizes equal to  $X_t - X_{t-}$ . Consider, for example, (1.1) with a two-dimensional drift-jump parameter  $\alpha$ , and a one-dimensional diffusion parameter  $\beta$ . For this model, an approximate martingale estimation function is defined by a function  $g = (g_\alpha, g_\beta)$ , where  $g_\alpha$  represents two coordinate functions related to the estimation of  $\alpha$ , and  $g_\beta$  one coordinate function associated with  $\beta$ . If  $g$  satisfies our conditions for rate optimal estimation of the diffusion parameter, then  $g_\beta(0, X_t, X_{t-}, \theta) = 0$  when  $t$  is a jump time. In other words, asymptotically, jumps in the data are not used for the estimation of  $\beta$ . When applied to continuous parts of the trajectory of the jump diffusion,  $g_\beta$  takes on the form used to define a rate optimal approximate martingale estimating function for the diffusion parameter of the corresponding continuous diffusion. When  $g$  also satisfies the conditions ensuring efficient estimation, then, asymptotically,  $g_\alpha$  too distinguishes perfectly between

jumps and continuous parts of the trajectory of the process. Specifically, for a jump time  $t$ ,  $g_\alpha(0, X_t, X_{t-}, \theta)$  has the form of the jump-related term in the score function of the continuously sampled jump-diffusion process; see [56]. At non-jump times,  $g$  has the structure ensuring an efficient estimating function for the drift and diffusion parameters of the corresponding continuous diffusion.

In this paper, we extend the established framework of high-frequency asymptotics for approximate martingale estimating functions for continuous diffusions to include jump-diffusion models. In particular, we maintain mathematically appealing assumptions regarding smoothness of the estimating functions. Under these assumptions, it is straightforward to obtain consistent, asymptotically normal estimators of the parameters of the jump-diffusions. The conditions for rate optimality and efficiency, however, impose further (inadvertently strict) restrictions, in terms of which models allow these conditions to be satisfied simultaneously with the smoothness assumptions on the estimating functions. Nonetheless, the conditions are enlightening. Taking into consideration the proofs used to arrive at our results, it is hard to imagine asymptotically well-performing estimators for jump-diffusions, which do not, essentially, conform to the conditions outlined above. Our findings indicate a path for future research in a framework where the estimating function is not required to be an approximate martingale estimating function satisfying the usual smoothness conditions. It seems likely that in conjunction with appropriate jump filtering, the extensive class of rate optimal and efficient approximate martingale estimating functions for continuous diffusions might be used to achieve rate optimal and efficient estimation of drift and diffusion parameters of jump-diffusions too.

The structure of the paper is as follows: Section 2 presents definitions, notation, and terminology used throughout the paper, as well as the main assumptions imposed on the jump-diffusion model and estimating functions. Section 3 presents the general theorem on the existence and asymptotics of consistent estimators based on approximate martingale estimating functions. Section 4 is devoted to criteria for rate optimality and efficiency of estimators of drift-jump and diffusion parameters. Section 5 contains central lemmas used to prove the main theorems, the proofs of these theorems, and the proofs of the lemmas. Appendix A consists of technical auxiliary results used in the proofs of the aforementioned lemmas.

## 2 Preliminaries

In this section we introduce basic notation, definitions, and regularity assumptions. Transposition of a matrix  $M$  is denoted by  $M^\star$ , and  $\|M\|$  represents the Euclidean norm. We denote by  $I_p$  the  $p \times p$  identity matrix. For any  $\mathbb{R}^p$ -valued function  $f$ , let  $f = (f_1, \dots, f_p)^\star$ , where  $f_j$  denotes the  $j$ th coordinate function. For an  $\mathbb{R}^q$ -valued argument  $u$ , let  $\partial_{u_k} f_j$  denote the partial derivative of  $f_j$  with respect to  $u_k$  and be the  $jk$ th element of the  $p \times q$  matrix  $\partial_u f$ . Let  $f^2 = (f_1^2, \dots, f_d^2)^\star$ . For a  $p \times q$

matrix-valued function  $F = (F_{jk})$ , we define  $\partial_u F = (\partial_u F_{jk})$  for real-valued  $u$  and  $F^2 = (F_{jk}^2)$ .

Let  $\Delta_0 = \max\{\Delta_n : n \in \mathbb{N}\}$ . Generic, strictly positive, real-valued constants are denoted by  $C$ . These constants may have dependencies emphasised by subscripts, and may also depend, implicitly, on e.g.  $\theta_0$ ,  $\Delta_0$ , and  $d$ , but never on the sample size  $n$ . Choose  $\varepsilon_0 > 0$  and let  $(0, \Delta_0)_{\varepsilon_0} = (0 - \varepsilon_0, \Delta_0 + \varepsilon_0)$ . A function  $f : (0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \rightarrow \mathbb{R}$  is said to be of polynomial growth in  $x$  and  $y$ , uniformly for  $t \in (0, \Delta_0)_{\varepsilon_0}$  and  $\theta$  in compact, convex sets if the following holds: For each compact, convex set  $K \subseteq \Theta$ , there exist constants  $C_K > 0$  such that for all  $x, y \in \mathcal{X}$ ,

$$\sup_{t \in (0, \Delta_0)_{\varepsilon_0}, \theta \in K} |f(t, y, x, \theta)| \leq C_K (1 + |x|^{C_K} + |y|^{C_K}) ,$$

or, equivalently,

$$\sup_{t \in (0, \Delta_0)_{\varepsilon_0}, \theta \in K} |f(t, y, x, \theta)| \leq C_K (1 + |x|^{C_K}) (1 + |y|^{C_K}) .$$

We use  $R(t, y, x, \theta)$  to denote generic functions defined on  $(0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta$ , which have coordinate functions of polynomial growth in  $x$  and  $y$ , uniformly for  $t \in (0, \Delta_0)_{\varepsilon_0}$  and  $\theta$  in compact, convex sets. In the same manner as  $C$ ,  $R$  may have both explicit and implicit dependencies. Functions  $R(t, x, \theta)$ ,  $R(y, x, \theta)$ , and  $R(t, x)$  are defined correspondingly.

**Definition 2.1.** We denote by  $\mathcal{C}_{\text{pol}}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$  the class of real-valued functions  $f(t, y, x, \theta) \in \mathcal{C}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$  which satisfy that  $f$  and its partial derivatives of all orders are of polynomial growth in  $x$  and  $y$ , uniformly for  $t \in (0, \Delta_0)_{\varepsilon_0}$  and  $\theta$  in compact, convex sets. The classes  $\mathcal{C}_{\text{pol}}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X} \times \Theta)$ ,  $\mathcal{C}_{\text{pol}}^\infty(\mathcal{X}^2 \times \Theta)$ ,  $\mathcal{C}_{\text{pol}}^\infty(\mathcal{X} \times \mathbb{R} \times \Theta)$ ,  $\mathcal{C}_{\text{pol}}^\infty(\mathcal{X} \times \Theta)$ , and  $\mathcal{C}_{\text{pol}}^\infty(\mathcal{X})$  are defined analogously for functions of the form  $f(t, x, \theta)$ ,  $f(y, x, \theta)$ ,  $f(y, \theta)$ , and  $f(y)$ .  $\diamond$

## 2.1 Model

Consider the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ , equipped with the  $(\mathcal{F}_t)_{t \geq 0}$ -adapted standard Wiener process  $\mathbf{W} = (W_t)_{t \geq 0}$  and the independent, time-homogeneous Poisson random measure  $N^\theta(dt, dz)$  on  $[0, \infty) \times \mathbb{R}$  with intensity measure  $\mu_\theta(dt, dz) = \nu_\theta(dz) dt$ . Here,  $\nu_\theta$  is a Lévy measure on  $\mathbb{R}$  with  $\nu_\theta(\{0\}) = 0$  and  $\nu_\theta(\mathbb{R}) < \infty$ . Let  $U^\theta$  be an  $\mathcal{F}_0$ -measurable random variable which is independent of  $\mathbf{W}$  and  $N^\theta$ . The process  $\mathbf{X}^\theta$  is assumed to solve (1.1) with the initial condition  $X_0^\theta = U^\theta$ . The drift, diffusion, and jump coefficients of the SDE,  $a, b : \mathcal{X} \times \Theta \rightarrow \mathbb{R}$  and  $c : \mathcal{X} \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ , respectively, are assumed to be known, deterministic functions. We make the following assumptions, among other reasons, in order to ensure that  $\mathbf{X}$  may be viewed as a càdlàg,  $(\mathcal{F}_t)$ -adapted Markov process.

**Assumption 2.2.** Suppose that

$$a(y, \theta), b(y, \theta) \in \mathcal{C}_{\text{pol}}^\infty(\mathcal{X} \times \Theta) \quad \text{and} \quad c(y, z, \theta) \in \mathcal{C}_{\text{pol}}^\infty(\mathcal{X} \times \mathbb{R} \times \Theta) .$$

Furthermore, the following holds for all  $\theta \in \Theta$ .

(i) For all  $y \in \mathcal{X}$ ,  $b^2(y, \theta) > 0$ .

(ii) There exist real-valued constants  $C_\theta > 0$  such that for all  $x, y \in \mathcal{X}$  and  $z \in \mathbb{R}$ ,

$$|a(x, \theta) - a(y, \theta)| + |b(x, \theta) - b(y, \theta)| \leq C_\theta |x - y|$$

and

$$|c(x, z, \theta) - c(y, z, \theta)| \leq C_\theta |x - y| (1 + |z|^{C_\theta}).$$

(iii) For all  $m \in \mathbb{N}$ ,

$$\sup_{t \in [0, \infty)} \mathbb{E}(|X_t^\theta|^m) < \infty.$$

(iv)  $\mathbf{X}^\theta$  is ergodic. That is, there exists an invariant probability measure  $\pi_\theta$ , such that for any  $\pi_\theta$ -integrable function  $f$ ,

$$\frac{1}{T} \int_0^T f(X_t^\theta) dt \xrightarrow{\mathcal{P}} \int_{\mathcal{X}} f(x) \pi_\theta(dx) \quad (2.1)$$

as  $T \rightarrow \infty$ . The measure  $\pi_\theta$  has moments of all orders.

(v) The Lévy measure  $\nu_\theta$  has density  $q(z, \theta) = \xi(\theta)p(z, \theta)$  with respect to a  $\sigma$ -finite measure  $\tilde{\nu}$ , where  $p(z, \theta)$  is a probability density with respect to  $\tilde{\nu}$ .

Finally, for the density of the Lévy measure:

(vi) It holds that  $(\theta \mapsto q(z, \theta)) \in \mathcal{C}^2(\Theta)$ . For each compact, convex set  $K \subseteq \Theta$ , there exists  $\varphi_K : \mathbb{R} \rightarrow [0, \infty)$  measurable with

$$\int_{\mathbb{R}} |z|^m \varphi_K(z) \tilde{\nu}(dz) < \infty$$

for all  $m \in \mathbb{N}_0$ , such that for all  $z \in \mathbb{R}$  and  $\theta \in K$ ,

$$q(z, \theta) + \sum_{j=1}^2 \sum_{k=1}^d |\partial_{\theta_k}^j q(z, \theta)| \leq \varphi_K(z).$$

◇

In the following, we put  $\pi_{\theta_0} = \pi$ . Note that by Assumption 2.2.(vi),  $\nu_\theta$  has moments of all orders for all  $\theta \in \Theta$ . Assumption 2.2 is similar to assumptions of e.g. [45, 48, 55]. See [42, 43] for conditions ensuring that an ergodic theorem of the form (2.1) holds, and under which  $\mathbf{X}$  has bounded moments as in Assumption 2.2.(iii).

Assuming ergodicity ensures the following lemma, which can be proved in the same way as the non-uniform part of [29, Lemma 8], using the Cauchy-Schwarz and Jensen's inequalities, and a version of [55, Proposition 3.1] (see also [45, p. 1598]).



**Lemma 2.3.** *Suppose that Assumption 2.2 holds, and that for fixed  $\theta \in \Theta$ , the functions  $x \mapsto f(x, \theta)$  and  $x \mapsto \partial_x f(x, \theta)$  are continuous and of polynomial growth in  $x$  for  $x \in \mathcal{X}$ . Then, pointwise for  $\theta \in \Theta$ ,*

$$\frac{1}{n} \sum_{i=1}^n f(X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} \int_{\mathcal{X}} f(x, \theta) \pi(dx).$$

◇

Now suppose that Assumption 2.2 holds, and let  $\lambda \in \Theta$ . Assume that  $f(t, y, x, \theta)$ ,  $f : ((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta) \rightarrow \mathbb{R}$ , and its partial derivatives  $\partial_y^i f$ ,  $i = 1, 2$ , exist, are continuous, and are of polynomial growth in  $x$  and  $y$ , uniformly for  $t \in (0, \Delta_0)_{\varepsilon_0}$  and  $\theta$  in compact, convex sets. Then, the infinitesimal generator  $\mathcal{L}_\lambda$  is defined by

$$\begin{aligned} \mathcal{L}_\lambda f(t, y, x, \theta) &= a(y, \lambda) \partial_y f(t, y, x, \theta) + \frac{1}{2} b^2(y, \lambda) \partial_y^2 f(t, y, x, \theta) \\ &\quad + \int_{\mathbb{R}} (f(t, y + c(y, z, \lambda), x, \theta) - f(t, y, x, \theta)) \nu_\lambda(dz). \end{aligned} \quad (2.2)$$

Often, the notation  $\mathcal{L}_\lambda f(t, y, x, \theta) = \mathcal{L}_\lambda(f(t, \theta))(y, x)$  is used, and we put  $\mathcal{L}_{\theta_0} = \mathcal{L}$ . Since  $\nu_\lambda(\mathbb{R}) < \infty$ , there exist constants  $C_{\lambda, \theta} > 0$  such that for all  $t \in (0, \Delta_0)_{\varepsilon_0}$ ,  $x, y \in \mathcal{X}$ , and  $\theta \in \Theta$ ,

$$\int_{\mathbb{R}} |f(t, y + c(y, z, \lambda), x, \theta) - f(t, y, x, \theta)| \nu_\lambda(dz) \leq C_{\lambda, \theta} (1 + |x|^{C_{\lambda, \theta}} + |y|^{C_{\lambda, \theta}}),$$

ensuring, in particular, that (2.2) is well-defined. More generally, it may be verified that integrals of the form  $\int f(\cdot, z) \nu_\lambda(dz)$  inherit polynomial growth properties of  $f$ . The operator  $\mathcal{L}_\lambda$ , always acting on the variable  $y$ , is defined correspondingly for e.g. functions of the form  $f(y)$ ,  $f(y, x, \theta)$ , and  $f(t, y, x, \mathbf{z}_k, \theta)$  where  $\mathbf{z}_k \in \mathbb{R}^k$ . In the latter case, the notation  $\mathcal{L}_\lambda f(t, y, x, \mathbf{z}_k, \theta) = \mathcal{L}_\lambda(f(t, \mathbf{z}_k, \theta))(y, x)$  is used.

We let  $\mathcal{L}_\lambda^k f = \mathcal{L}_\lambda(\mathcal{L}_\lambda^{k-1} f)$  for  $k \in \mathbb{N}$  with  $\mathcal{L}_\lambda^0 f = f$ . When  $f$  is  $\mathbb{R}^d$ -valued, and the generator is well-defined for each coordinate function, then  $\mathcal{L}_\lambda f = (\mathcal{L}_\lambda f_1, \dots, \mathcal{L}_\lambda f_d)^*$ . If  $F$  is a matrix-valued function,  $\mathcal{L}_\lambda F$  denotes the matrix with  $ij$ th element  $\mathcal{L}_\lambda F_{ij}$ .

The infinitesimal generator notation is useful for expressing the following lemma.

**Lemma 2.4.** *Let Assumption 2.2 hold. For some  $k \in \mathbb{N}_0$ , suppose that  $f(y, x, \theta)$ ,  $f : \mathcal{X}^2 \times \Theta \rightarrow \mathbb{R}$ , and its partial derivatives  $\partial_y^i f$ ,  $i = 1, \dots, 2(k+1)$ , exist, are continuous, and are of polynomial growth in  $x$  and  $y$ , uniformly for  $\theta$  in compact, convex sets. Then, for  $0 \leq t < t + \Delta$ ,  $\Delta \leq \Delta_0$ , and  $\lambda \in \Theta$ ,*

$$\mathbb{E} \left( f(X_{t+\Delta}^\lambda, X_t^\lambda, \theta) \mid X_t^\lambda \right) = \sum_{i=0}^k \frac{\Delta^i}{i!} \mathcal{L}_\lambda^i f(X_t^\lambda, X_t^\lambda, \theta) + \Delta^{k+1} R_\lambda(\Delta, X_t^\lambda, \theta),$$

where

$$\Delta^{k+1} R_\lambda(\Delta, X_t^\lambda, \theta) =$$

$$\int_0^{\Delta_0} \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E} \left( \mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}^\lambda, X_t^\lambda, \theta) \mid X_t^\lambda \right) du_{k+1} \cdots du_2 du_1.$$

◇

Lemma 2.4 is effectively a jump-diffusion extension of the expression given by e.g. [9, Lemma 1] for continuous diffusions. Notationally, the proof of Lemma 2.4 is very similar to the proof of the continuous version, see [57, Lemma 1.10] and [29, Lemma 1], but it uses Itô's formula for diffusions with jumps and the infinitesimal generator (2.2). The lemma is useful for verifying the approximate martingale property (2.3), and for creating approximate martingale estimating functions. It is also key to proving Lemma 2.7, which emphasises two important properties of the estimating functions considered here.

## 2.2 Estimating Functions

**Definition 2.5.** Let  $g : (0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \rightarrow \mathbb{R}^d$ . Suppose there exists a constant  $\kappa_0 \geq 2$ , such that for all  $n \in \mathbb{N}$ ,  $i = 1, \dots, n$ , and  $\theta \in \Theta$ ,

$$\mathbb{E} \left( g(\Delta_n, X_{n,i}^\theta, X_{n,i-1}^\theta, \theta) \mid X_{n,i-1}^\theta \right) = \Delta_n^{\kappa_0} R_\theta(\Delta_n, X_{n,i-1}^\theta). \quad (2.3)$$

Then, (1.2) defines an approximate martingale estimating function. ◇

A  $G_n$ -estimator  $\hat{\theta}_n$  is obtained as a solution to the estimating equation  $G_n(\theta) = 0$ , see also [25, Definition 2.3]. When (2.3) is satisfied with  $R_\theta(t, x)$  identically equal to zero, (1.2) is referred to as a martingale estimating function.

Let  $M_n$  be any invertible  $d \times d$  matrix with real entries which may depend on, e.g.,  $\Delta_n$  and  $\theta_0$ . Then,  $G_n(\theta)$  and  $M_n G_n(\theta)$  produce identical estimators of  $\theta$ . These estimating functions are considered versions of each other. It is enough that one version satisfies the assumptions set forth in this paper.

We make the following assumptions about the function  $g$ , which defines our estimating function (1.2).

**Assumption 2.6.** Choose some  $\varepsilon_0 > 0$ .

(i) The function  $g(t, y, x, \theta)$  satisfies Definition 2.5 for some  $\kappa_0 \geq 2$ .

(ii) For  $j = 1, \dots, d$ , it holds that

$$g_j(t, y, x, \theta) \in C_{pol}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta).$$

(iii) For  $k = 0, 1$ , and all  $t \in (0, \Delta_0)_{\varepsilon_0}$ ,  $x, y \in \mathcal{X}$ , and  $\theta \in \Theta$ , the expansion

$$\begin{aligned} & \partial_\theta^k g(t, y, x, \theta) \\ &= \partial_\theta^k g(0, y, x, \theta) + t \partial_\theta^k g^{(1)}(y, x, \theta) + \frac{1}{2} t^2 \partial_\theta^k g^{(2)}(y, x, \theta) + t^3 R(t, y, x, \theta) \end{aligned} \quad (2.4)$$

holds, where  $g_j^{(i)}(y, x, \theta) = \partial_t^i g_j(0, y, x, \theta)$ . ◇

In this paper, the assumptions of  $\mathcal{C}^\infty$ -smoothness and polynomial growth, together with assumptions on the moments of, e.g.,  $X_t^\theta$ ,  $\nu_\theta$ , and  $\pi_\theta$ , serve to simplify the exposition and proofs, and could be relaxed; see also [24, Chapter 3].

For use in the proofs, we note the following lemma.

**Lemma 2.7.** *Suppose that Assumptions 2.2 and 2.6 hold. Then, for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ ,*

$$g(0, x, x, \theta) = 0 \quad \text{and} \quad g^{(1)}(x, x, \theta) = -\mathcal{L}_\theta(g(0, \theta))(x, x).$$

◇

Lemma 2.7 is similar to [58, Lemma 2.3], to which we refer for a proof.

### 3 General asymptotic theory

This section presents the general theorem on the existence and properties of consistent estimators of  $\theta$  in the model (1.1), based on approximate martingale estimating functions.

**Assumption 3.1.** *Let*

$$\begin{aligned} A(\lambda, \theta) &= \int_{\mathcal{X}} (\mathcal{L}_\theta(g(0, \lambda))(x, x) - \mathcal{L}_\lambda(g(0, \lambda))(x, x)) \pi_\theta(dx) \\ B(\theta, \theta) &= \int_{\mathcal{X}} (\mathcal{L}_\theta(\partial_\theta g(0, \theta))(x, x) - \partial_\theta \mathcal{L}_\theta(g(0, \theta))(x, x)) \pi_\theta(dx) \\ C(\theta, \theta) &= \int_{\mathcal{X}} \mathcal{L}_\theta(gg^*(0, \theta))(x, x) \pi_\theta(dx). \end{aligned}$$

*The following holds for all  $\theta \in \Theta$ .*

- (i) *The  $\mathbb{R}^d$ -vector  $A(\lambda, \theta)$  is non-zero whenever  $\lambda \neq \theta$ .*
- (ii) *The  $d \times d$  matrix  $B(\theta, \theta)$  is non-singular.*
- (iii) *The symmetric  $d \times d$  matrix  $C(\theta, \theta)$  is positive definite.*

◇

**Theorem 3.2.** *Consider the model given by (1.1). Suppose that Assumptions 2.2, 2.6, and 3.1 hold. If (2.3) holds with  $R_\theta(t, x)$  not identically equal to 0, suppose also that  $n\Delta_n^{2\kappa_0-1} \rightarrow 0$  as  $n \rightarrow \infty$ . Then,*

- (i) *there exists a consistent  $G_n$ -estimator  $\hat{\theta}_n$ . Choose any compact, convex set  $K \subseteq \Theta$  with  $\theta_0 \in \text{int}K$ , where  $\text{int}K$  denotes the interior of  $K$ . Then,  $\hat{\theta}_n$  is eventually unique in  $K$ , in the sense that for any  $G_n$ -estimator  $\tilde{\theta}_n$  with  $\mathbb{P}(\tilde{\theta}_n \in K) \rightarrow 1$  as  $n \rightarrow \infty$ , it holds that  $\mathbb{P}(\hat{\theta}_n \neq \tilde{\theta}_n) \rightarrow 0$  as  $n \rightarrow \infty$ .*

(ii) for any consistent  $G_n$ -estimator  $\hat{\theta}_n$ , it holds that

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V(\theta_0)),$$

where  $V(\theta_0) = B(\theta_0, \theta_0)^{-1}C(\theta_0, \theta_0)(B(\theta_0, \theta_0)^*)^{-1}$  is positive definite.

(iii) for any consistent  $G_n$ -estimator  $\hat{\theta}_n$ ,

$$\begin{aligned} \hat{V}_n &= n\Delta_n \left( \sum_{i=1}^n \partial_{\theta} g(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right)^{-1} \left( \sum_{i=1}^n g g^*(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right) \\ &\quad \times \left( \sum_{i=1}^n \partial_{\theta} g^*(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right)^{-1} \end{aligned}$$

is a consistent estimator of  $V(\theta_0)$ , so

$$\sqrt{n\Delta_n} \hat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I_d),$$

where  $\hat{V}_n^{1/2}$  is the unique, positive semidefinite square root of  $\hat{V}_n$ .

◇

The proof of Theorem 3.2 is given in Section 5.2. The condition  $n\Delta_n^{2\kappa_0-1} \rightarrow 0$  is necessary for the asymptotic normality in (ii) and (iii), when the estimating function is not an exact martingale. These results follow from the asymptotic normality of the normalized estimating function, see Lemma 5.1, which is proved using a central limit theorem for martingales. The normalized estimating function differs from a martingale by a term of order  $(n\Delta_n^{2\kappa_0-1})^{1/2}$ , which must necessarily vanish. In case  $n\Delta_n^{2\kappa_0-1}$  converges to a constant different from zero, the asymptotic normal distribution has a mean different from zero.

*Example 3.3.* Consider the SDE

$$dX_t^\theta = \tilde{a}(X_t^\theta, \alpha) dt + \tilde{b}(X_t^\theta, \beta) \sigma dW_t + \int_{\mathbb{R}} \tilde{b}(X_{t-}^\theta, \beta) z (N - \mu)(dt, dz).$$

The drift parameter  $\alpha$ , and the diffusion-jump parameter  $\beta$  are the unknown parameters to be estimated. Note that the Poisson random measure  $N$  and its intensity measure  $\mu$  do not depend on the parameters. For simplicity, let  $\alpha \in A \subseteq \mathbb{R}$  and  $\beta \in B \subseteq \mathbb{R}$  so that  $d = 2$  (the results generalise to larger  $d$  as well). Put  $\theta = (\alpha, \beta)^*$  and  $\Theta = A \times B$ , and suppose that Assumption 2.2 holds. Furthermore, suppose that  $\sigma^2 + \gamma_2 = 1$ , where  $\gamma_k$  denotes the  $k$ th moment of the Lévy measure  $\nu$ .

By Lemma 2.4, for  $\theta \in \Theta$  and  $0 \leq t < t + \Delta$  with  $\Delta \leq \Delta_0$ ,

$$\begin{aligned} \mathbb{E}(X_{t+\Delta}^\theta \mid X_t^\theta) &= X_t^\theta + \Delta \tilde{a}(X_t^\theta, \alpha) + \Delta^2 R_\theta(\Delta, X_t^\theta) \\ \mathbb{E}((X_{t+\Delta}^\theta - X_t^\theta)^2 \mid X_t^\theta) &= \Delta \tilde{b}^2(X_t^\theta, \beta) + \Delta^2 R_\theta(\Delta, X_t^\theta) \end{aligned}$$

so, under weak conditions on the functions  $m_1(x, \theta)$  and  $m_2(x, \theta)$ ,

$$g(t, y, x, \theta) = \begin{pmatrix} m_1(x, \theta) (y - x - t\tilde{a}(x, \alpha)) \\ m_2(x, \theta) \left( (y - x - t\tilde{a}(x, \alpha))^2 - t\tilde{b}^2(x, \beta) \right) \end{pmatrix}$$

satisfies Assumption 2.6 with  $\kappa_0 = 2$ .

Suppose also that Assumption 3.1 holds, and that  $n\Delta_n^3 \rightarrow 0$  as  $n \rightarrow \infty$ . Then, by Theorem 3.2.(ii), for any consistent  $G_n$ -estimator  $\hat{\theta}_n$ ,

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_2(0, V(\theta_0)), \quad (3.1)$$

where  $V(\theta_0) = B(\theta_0, \theta_0)^{-1}C(\theta_0, \theta_0)(B(\theta_0, \theta_0)^*)^{-1}$  with

$$B(\theta_0, \theta_0) = - \int_{\mathcal{X}} \begin{pmatrix} m_1(x, \theta_0) \partial_{\alpha} \tilde{a}(x, \alpha_0) & 0 \\ 0 & m_2(x, \theta_0) \partial_{\beta} \tilde{b}^2(x, \beta_0) \end{pmatrix} \pi(dx)$$

and

$$C(\theta_0, \theta_0) = \int_{\mathcal{X}} \begin{pmatrix} m_1^2(x, \theta_0) \tilde{b}^2(x, \beta_0) & m_1 m_2(x, \theta_0) \tilde{b}^3(x, \beta_0) \gamma_3 \\ m_2 m_1(x, \theta_0) \tilde{b}^3(x, \beta_0) \gamma_3 & m_2^2(x, \theta_0) \tilde{b}^4(x, \beta_0) \gamma_4 \end{pmatrix} \pi(dx).$$

◦

The SDE and the estimating function used in Example 3.3 correspond to those considered in [44], incorporated into the current parametric framework. The result (3.1) is in accordance with [44, Theorem 3.4]. Similarly, in the particular case of quadratic approximate martingale estimating functions, the result in Theorem 3.2.(ii) essentially follows from [45, Theorem 2.9] by interpreting the quasi-likelihood estimator proposed in [45] as a quadratic approximate martingale estimating function.

## 4 Rate optimality and efficiency

Here, we investigate conditions ensuring rate-optimal and efficient estimators. In Section 4.1, we discuss the optimal rate and Fisher information for jump-diffusion models. In Sections 4.2 and 4.3, conditions are given, which are designed to ensure rate optimality and efficiency of  $G_n$ -estimators in two specific types of submodels of (1.1). The interpretation and implications of these conditions are discussed.

Suppose in the following that  $A \subseteq \mathbb{R}^{d_1}$  and  $B \subseteq \mathbb{R}^{d_2}$  with  $d = d_1 + d_2$ , and put  $\Theta = A \times B$ . Consider, for  $\alpha \in A$ ,  $\beta \in B$ , and  $\theta = (\alpha, \beta)$ , the SDE

$$dX_t^{\theta} = a(X_t^{\theta}, \alpha) dt + b(X_t^{\theta}, \beta) dW_t + \int_{\mathbb{R}} c(X_{t-}^{\theta}, z, \alpha) N^{\alpha}(dt, dz), \quad (4.1)$$

where  $N^{\alpha}$  has intensity measure  $\mu_{\alpha}(dt, dz) = \nu_{\alpha}(dz) dt$ . The parameters  $\alpha$  and  $\beta$  are referred to as the drift-jump and diffusion parameters, respectively. For simplicity, the following assumption is introduced.

**Assumption 4.1.** Let  $c_{x,\alpha}(z) = c(x, z, \alpha)$ . One of the following cases (a) or (b) applies for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ .

- (a) The dominating measure  $\tilde{\nu}$  of the family of Lévy measures is Lebesgue measure. The set  $\mathcal{W}(x) = c_{x,\alpha}(\mathbb{R})$  is open and does not depend on  $\alpha$ . The mapping  $z \mapsto c_{x,\alpha}(z)$  is bijective with a continuously differentiable inverse  $w \mapsto c_{x,\alpha}^{-1}(w)$ . Let

$$\varphi(x, w, \alpha) = q(c_{x,\alpha}^{-1}(w), \alpha) |\partial_w c_{x,\alpha}^{-1}(w)|, \quad w \in \mathcal{W}(x),$$

be the transformation of the Lévy density  $q(\cdot, \alpha)$  by  $z \mapsto c_{x,\alpha}(z)$ , and let  $\eta_x$  denote Lebesgue measure on  $\mathcal{W}(x)$ .

- (b) The dominating measure  $\tilde{\nu}$  is counting measure on an at most countable set  $\mathcal{Q} \subset \mathbb{R}$ , and  $c_{x,\alpha}(z) = c_x(z)$  for all  $z \in \mathcal{Q}$ . Define  $\mathcal{W}(x) = c_x(\mathcal{Q})$  and

$$\varphi(x, w, \alpha) = \sum_{z \in c_x^{-1}(\{w\})} q(z, \alpha),$$

and let  $\eta_x$  denote counting measure on  $\mathcal{W}(x)$ .

In both cases, for all  $x \in \mathcal{X}$ ,

$$\partial_\theta \int_{\mathcal{W}(x)} g(0, x + w, x, \theta) \varphi(x, w, \alpha) \eta_x(dw) = \int_{\mathcal{W}(x)} \partial_\theta (g(0, x + w, x, \theta) \varphi(x, w, \alpha)) \eta_x(dw).$$

◇

#### 4.1 Conjecture on rate optimality and efficiency

Local asymptotic normality has not yet been established for the general model (4.1) under the present observation scheme and asymptotic scenario. However, when  $\mathbf{X}$  is ergodic, and under Assumption 4.1 and suitable regularity conditions, we conjecture the following: The jump-diffusion model is locally asymptotically normal with rate  $\sqrt{n\Delta_n}$  for the drift-jump parameter  $\alpha$ , rate  $\sqrt{n}$  for the diffusion parameter  $\beta$ , and Fisher information

$$\mathcal{I}(\theta) = \begin{pmatrix} \mathcal{I}_1(\theta) & 0 \\ 0 & \mathcal{I}_2(\theta) \end{pmatrix}, \quad (4.2)$$

where

$$\mathcal{I}_1(\theta) = \int_{\mathcal{X}} \left( \frac{\partial_\alpha a(x, \alpha)^* \partial_\alpha a(x, \alpha)}{b^2(x, \beta)} \right) + \int_{\mathcal{W}(x)} \frac{\partial_\alpha \varphi(x, w, \alpha)^* \partial_\alpha \varphi(x, w, \alpha)}{\varphi(x, w, \alpha)} \eta_x(dw) \pi_\theta(dx) \quad (4.3)$$

and

$$\mathcal{I}_2(\theta) = \int_{\mathcal{X}} \frac{\partial_\beta b^2(x, \beta)^* \partial_\beta b^2(x, \beta)}{2b^4(x, \beta)} \pi_\theta(dx).$$

The measure  $\eta_x$  is either Lebesgue measure or counting measure on  $\mathcal{W}(x)$ , see Assumption 4.1. In particular, we conjecture that a consistent estimator  $\hat{\theta}_n^* = (\hat{\alpha}_n^*, \hat{\beta}_n^*)$  of  $\theta^* = (\alpha^*, \beta^*)$  is rate optimal if

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} Z, \quad (4.4)$$

where  $Z$  is a zero-mean,  $d$ -dimensional random vector with a positive definite covariance matrix, and that  $\hat{\theta}_n$  is efficient if

$$Z \sim \mathcal{N}_d(0, \mathcal{I}(\theta_0)^{-1}). \quad (4.5)$$

The conjecture is motivated by the following observations. Here, the results quoted from the literature are given in a generality suitable for our framework and, to keep the discussion concise, regularity conditions are not included. Let  $\widehat{\mathbf{X}}_T = (X_t)_{0 \leq t \leq T}$  denote continuous-time observations of the sample path of  $\mathbf{X}$  over the interval  $[0, T]$  for  $T > 0$ , and let  $\widehat{X}_n = (X_0^n, X_1^n, \dots, X_n^n)$  denote discrete-time observations of  $\mathbf{X}$  sampled as in this paper.

For continuous diffusions, local asymptotic normality for discrete-time observations  $\widehat{X}_n$  with rate  $\sqrt{n\Delta_n}$  for  $\alpha$ ,  $\sqrt{n}$  for  $\beta$ , and Fisher information (4.2) was shown in [14, Theorem 4.1]. There is no reason to expect it to be possible to estimate the diffusion parameter at a faster rate or more accurately in the jump-diffusion framework considered here.

In [56], likelihood theory was developed for continuous-time data  $\widehat{\mathbf{X}}_T$  for models of the type (4.1) with only a drift-jump parameter  $\alpha$ , under the assumption that the diffusion coefficient  $b(x)$  is known. (This assumption is necessary for continuous time data.) The rate of convergence is  $\sqrt{T}$  and, in case (a) of Assumption 4.1, it is seen from formulas (3.4), (3.6), and Corollary 3.3 of [56] that the Fisher information is given by (4.3). There is no reason to believe that the drift-jump parameter can be estimated at a faster rate or more accurately from discrete-time data. In [13], the local asymptotic normality property was established for continuous-time data when  $c(x, z, \alpha) = \gamma(x)z$  and the Poisson random measure does not depend on  $\alpha$ .

For the general model (4.1), an estimator  $\hat{\theta}_n$  based on  $\widehat{X}_n$  was derived in [55], which, in case (a) of Assumption 4.1, satisfies (4.4) and (4.5), provided that  $n\Delta_n^2 \rightarrow \infty$ . Thus, the conjectured rate and Fisher information can be achieved. Comparison to the result in [14] indicates that the rate and Fisher information for the diffusion parameter must necessarily be as conjectured, while comparison to the result in [56] indicates that this is true of the drift-jump parameter too.

For certain submodels of (4.1), the following papers showed results on local asymptotic normality, which are in accordance with our conjecture. All results are for discrete-time data with infinite-horizon, high-frequency asymptotics. In [3, Propositions 1 & 2], the model (4.1) is investigated with  $b(x, \beta) \equiv b(x)$ ,  $c(x, z, \alpha) \equiv z$  in

case (a) of Assumption 4.1. It is assumed that  $n\Delta_n^2 \rightarrow 0$  as  $n \rightarrow \infty$ . The model studied in [3] is more general than our model in one respect: the jump intensity is allowed to be state-dependent. In [32, Theorem 2.2], the model (4.1) is studied with only a one-dimensional drift parameter  $\alpha$  in either of the cases (a) or (b) of Assumption 4.1. The diffusion coefficient and jump mechanism are assumed to be known. In [27], the Ornstein-Uhlenbeck process

$$dX_t = -\alpha_1(X_t - \alpha_2)dt + \beta dW_t + \int_{\mathbb{R}} z(N - \mu)(dt, dz)$$

is considered, where the random measure  $N$  and its intensity measure  $\mu$  do not depend on  $\alpha$ . Both (a) and (b) of Assumption 4.1 are covered. Furthermore, local asymptotic normality is shown in [60], when the random measure is given by a Poisson process with the intensity  $\lambda$  a parameter to be estimated. Finally, the model

$$dX_t = (\alpha - \gamma)dt + \beta dW_t + \int_{\mathbb{R}} zN(dt, dz)$$

is investigated in [31] with  $\nu_\gamma(dz) = \gamma\varepsilon_1(dz)$ , where  $\varepsilon_1$  is the degenerate probability measure with point mass in 1, and the unknown parameter  $\theta = (\alpha, \gamma, \beta)$  is three-dimensional.

## 4.2 Drift-jump parameter

In this section, we consider the submodel of (4.1) given by

$$dX_t^\alpha = a(X_t^\alpha, \alpha)dt + b(X_t^\alpha)dW_t + \int_{\mathbb{R}} c(X_{t-}^\alpha, z, \alpha)N^\alpha(dt, dz), \quad (4.6)$$

where  $\alpha \in A$ , and  $\Theta = A$  is a non-empty, open subset of  $\mathbb{R}^d$ . Theorem 3.2 already yields the conjectured optimal rate for estimators of the parameter  $\alpha$ . In order to ensure efficiency, we impose the following sufficient condition, which presupposes Assumption 4.1.

**Condition 4.2.** *For each  $\alpha \in A$ , there exists an invertible  $d \times d$  matrix  $K_\alpha$  such that for all  $x \in \mathcal{X}$ , and  $\eta_x$ -almost all  $w \in \mathcal{W}(x)$ ,*

$$\partial_y g(0, x, x, \alpha) = K_\alpha \frac{\partial_\alpha a(x, \alpha)^*}{b^2(x)} \quad \text{and} \quad g(0, x + w, x, \alpha) = K_\alpha \frac{\partial_\alpha \varphi(x, w, \alpha)^*}{\varphi(x, w, \alpha)}.$$

◇

Here,  $\partial_y g(0, x, x, \alpha)$  denotes  $\partial_y g(0, y, x, \alpha)$  evaluated in  $y = x$ . Using Remark 5.2, the following Corollary 4.3 may easily be verified.

**Corollary 4.3.** *Consider the model given by (4.6). Suppose that the assumptions of Theorem 3.2, as well as Assumption 4.1 and Condition 4.2 hold. Then, any consistent  $G_n$ -estimator  $\hat{\alpha}_n$  has asymptotic variance  $\mathcal{I}_1^{-1}(\alpha_0)$  given by (4.3).* ◇



Thus  $\hat{\alpha}_n$  is efficient, provided our conjecture is correct. The first equation in Condition 4.2 corresponds to the condition given in [58, Condition 1.2] for efficiency of estimators of the drift parameter of a continuous diffusion. The second equation is of a type not seen in [58]. It is a jump-related condition on the off-diagonal  $y \neq x$  of  $g(t, y, x, \theta)$  when  $t = 0$ . In the limit  $\Delta_n \rightarrow 0$ , the full sample path of  $\mathbf{X}$  is observed and  $g(0, y, x, \alpha)$  and its derivatives are evaluated in  $y = X_t$  and  $x = X_{t-}$ . For continuous diffusions,  $X_t = X_{t-}$  for all  $t$ , so conditions for  $y \neq x$  are not relevant. For jump-diffusions, however,  $X_t \neq X_{t-}$  whenever  $t$  is a jump time so, in this case, off-diagonal conditions are expected.

It is evident that an estimating function which satisfies Condition 4.2 discriminates, asymptotically, between situations where  $X_t = X_{t-}$  and  $X_t \neq X_{t-}$ . When  $X_t = X_{t-}$ , the function must behave like an efficient estimating function for the drift parameter of the corresponding continuous diffusion. However, when  $X_t \neq X_{t-}$ , asymptotically, the function takes on the form of the term associated with jumps in the score function of the continuously sampled jump-diffusion process [56]. In essence, the estimating function must, in the limit  $\Delta_n \rightarrow 0$ , equal the score function of the jumps at  $(y, x) = (X_t, X_{t-})$  when  $X_t \neq X_{t-}$ . This severely restricts the class of efficient estimators for jump-diffusions in contrast to continuous diffusions.

### 4.3 Drift-jump and diffusion parameter

In this section, we consider the model (4.1) where  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$  are non-empty, open sets. Let  $g_\alpha = (g_1, g_2)^*$  and  $g_\beta = g_3$ . For convenience, we introduce the following notation. Let  $\mathcal{S}_\alpha$  denote the support of the Lévy-measure  $\nu_\alpha$ , and put  $\tau_\alpha(y, z) = y + c(y, z, \alpha)$ . Define the sets  $\mathcal{M}_k(y, \alpha)$  iteratively by  $\mathcal{M}_0(y, \alpha) = \{y\}$ , and  $\mathcal{M}_k(y, \alpha) = \tau_\alpha(\mathcal{M}_{k-1}(y, \alpha), \mathcal{S}_\alpha)$  for  $k \in \mathbb{N}$ . The set  $\mathcal{M}_k(y, \alpha)$  is the subset of  $\mathcal{X}$  that can be reached from the point  $y$  by  $k$  pure jumps, when  $\alpha$  is the true drift-jump parameter. The following Condition 4.4 is an additional condition for use in Theorem 4.5. The theorem improves on Theorem 3.2.(ii), in that it yields the conjectured optimal rate of convergence for consistent  $G_n$ -estimators  $\hat{\theta}_n^* = (\hat{\alpha}_n^*, \hat{\beta}_n)$ .

**Condition 4.4.** *Suppose that for all  $\tilde{\alpha} \in A$ ,  $\theta \in \Theta$ , and  $x \in \mathcal{X}$*

$$\begin{aligned} g_\beta(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 1, 2, 3, 4 \\ \partial_y g_\beta(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1, 2, 3 \\ \partial_y^2 \partial_\alpha g_\beta(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1 \\ \partial_\alpha g_\beta^{(1)}(y, x, \theta) &= 0, & y \in \mathcal{M}_1(x, \tilde{\alpha}). & \end{aligned}$$

◇

**Theorem 4.5.** *Consider the model given by (4.1) with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that all the assumptions of Theorem 3.2 except Assumption 3.1.(iii) hold, and that*

Condition 4.4 holds. If (2.3) holds with  $R_\theta(t, x)$  not identically equal to 0, suppose also that  $n\Delta_n^{2(\kappa_0-1)} \rightarrow 0$  as  $n \rightarrow \infty$ . Let

$$\begin{aligned}
B_1(\theta_0, \theta_0) &= - \int_{\mathcal{X}} \partial_y g_\alpha(0, x, x, \theta_0) \partial_\alpha a(x, \alpha_0) \pi(dx) \\
&\quad - \int_{\mathcal{X}} \int_{\mathbb{R}} \partial_y g_\alpha(0, x + c(x, z, \alpha_0), x, \theta_0) \partial_\alpha c(x, z, \alpha_0) \nu_{\alpha_0}(dz) \pi(dx) \\
&\quad - \int_{\mathcal{X}} \int_{\mathbb{R}} g_\alpha(0, x + c(x, z, \alpha_0), x, \theta_0) \partial_\alpha q(z, \alpha_0) \tilde{\nu}(dz) \pi(dx), \\
B_2(\theta_0, \theta_0) &= - \int_{\mathcal{X}} \frac{1}{2} \partial_y^2 g_\beta(0, x, x, \theta_0) \partial_\beta b^2(x, \beta_0) \pi(dx), \\
D_1(\theta, \theta) &= \int_{\mathcal{X}} b^2(x, \beta) \partial_y g_\alpha(\partial_y g_\alpha)^*(0, x, x, \theta) \pi_\theta(dx) \\
&\quad + \int_{\mathcal{X}} \int_{\mathbb{R}} g_\alpha g_\alpha^*(0, x + c(x, z, \alpha), x, \theta) \nu_\alpha(dz) \pi_\theta(dx), \\
D_2(\theta, \theta) &= \int_{\mathcal{X}} \frac{1}{2} b^4(x, \beta) (\partial_y^2 g_\beta(0, x, x, \theta))^2 \pi_\theta(dx),
\end{aligned}$$

and assume that  $D_1(\theta, \theta)$  is invertible, and  $D_2(\theta, \theta) \neq 0$  for all  $\theta \in \Theta$ . Then, for any consistent  $G_n$ -estimator  $\hat{\theta}_n$ , it holds that

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, V(\theta_0)) \quad (4.7)$$

where

$$V(\theta_0) = \begin{pmatrix} B_1(\theta_0, \theta_0)^{-1} D_1(\theta_0, \theta_0) (B_1(\theta_0, \theta_0)^*)^{-1} & 0 \\ 0 & B_2(\theta_0, \theta_0)^{-2} D_2(\theta_0, \theta_0) \end{pmatrix}$$

is positive definite. Furthermore,  $\hat{V}_n = \text{diag}(\hat{V}_{n,1}, \hat{V}_{n,2})$  given by

$$\begin{aligned}
\hat{V}_{n,1} &= n\Delta_n \left( \sum_{i=1}^n \partial_\alpha g_\alpha(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right)^{-1} \left( \sum_{i=1}^n g_\alpha g_\alpha^*(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right) \\
&\quad \times \left( \sum_{i=1}^n \partial_\alpha g_\alpha^*(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right)^{-1} \\
\hat{V}_{n,2} &= n \left( \sum_{i=1}^n \partial_\beta g_\beta(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n) \right)^{-2} \sum_{i=1}^n g_\beta^2(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n)
\end{aligned}$$

is a consistent estimator of  $V(\theta_0)$ , so

$$\hat{V}_n^{-1/2} \begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, I_3), \quad (4.8)$$

where  $\hat{V}_n^{1/2}$  is the unique, positive semidefinite square root of  $\hat{V}_n$ .  $\diamond$

The proof of Theorem 4.5 is given in Section 5.2. When  $G_n(\theta)$  is not a martingale estimating function, the condition  $n\Delta_n^{2(\kappa_0-1)} \rightarrow 0$  is necessary for (4.7) and (4.8). Here,  $\Delta_n$  must converge to zero a bit faster than in Theorem 3.2. The reason is that, under the conditions ensuring rate optimality, the third coordinate of the estimating function must be normalized differently from the first two coordinates. The third coordinate of the difference between the normalized estimating function and the normalized martingale is of order  $(n\Delta_n^{2(\kappa_0-1)})^{1/2}$ , while the other two are of the same order as in Theorem 3.2.

Making use of Remark 5.2, it is evident that efficiency of the estimators according to our conjecture is ensured by the following additional Condition 4.6, which presupposes Assumption 4.1.

**Condition 4.6.** *For all  $\theta \in \Theta$  there exists an invertible  $2 \times 2$  matrix  $K_\theta^{(1)}$  and a constant  $K_\theta^{(2)} \neq 0$ , such that*

$$\partial_y g_\alpha(0, x, x, \theta) = K_\theta^{(1)} \frac{\partial_\alpha a(x, \alpha)^*}{b^2(x, \beta)}, \quad \partial_y^2 g_\beta(0, x, x, \theta) = K_\theta^{(2)} \frac{\partial_\beta b^2(x, \beta)}{b^4(x, \beta)},$$

and

$$g_\alpha(0, x + w, x, \theta) = K_\theta^{(1)} \frac{\partial_\alpha \varphi(x, w, \alpha)^*}{\varphi(x, w, \alpha)}$$

for  $\eta_x$ -almost all  $w \in \mathcal{W}(x)$ , and all  $x \in \mathcal{X}$ .  $\diamond$

**Corollary 4.7.** *Suppose that the assumptions of Theorem 4.5, as well as Assumption 4.1 and Condition 4.6 hold. Then, any consistent  $G_n$ -estimator  $\hat{\theta}_n$  has asymptotic variance  $\mathcal{I}^{-1}(\theta_0)$  given by (4.2).  $\diamond$*

Thus  $\hat{\theta}_n$  is efficient, provided our conjecture is correct. The additional condition for efficiency in Condition 4.6, compared to Condition 4.2, is identical to the one identified in [58] for the diffusion parameter of a continuous diffusion, and in [25] for the diffusion parameter of a continuous diffusion in the case of infill asymptotics.

Condition 4.4 requires that the  $g_\beta$  coordinate of  $g$  as well as several of its derivatives vanish at a number of points depending on the jump dynamics of the process. This reveals that for many SDE models, rate optimal and efficient estimation of the diffusion parameter is not feasible when using the smooth estimating functions considered in this paper. In Theorem 4.5, the non-degeneracy condition on  $D_2(\theta, \theta)$  requires that  $\partial_y^2 g_\beta(0, x, x, \theta)$  does not vanish  $\pi_\theta$ -almost surely for any  $\theta$ . This easily conflicts with Condition 4.4. For example, for models where the first equation of the condition amounts to the requirement that  $g_\beta(0, y, x, \beta) = 0$  for all  $x, y \in \mathcal{X}$ , it is clear that the non-degeneracy condition cannot be satisfied.

Finally, it may be noted that for special cases of (4.1) with only a one-dimensional diffusion parameter and no drift-jump parameter, Condition 4.4 may be reduced

to its first and second equations involving  $g_\beta$  and  $\partial_y g_\beta$  for  $k = 1, 2$  and  $k = 0, 1$  respectively, see [24, Section 3.4.4] for further details.

For continuous diffusions, conditions under which an approximate martingale estimating function is rate optimal and efficient are quite straightforward, and it is easy to find estimating functions which satisfy the conditions. This was concluded in [58] for the current sampling scheme, and in [25] for infill asymptotics. The present paper demonstrates that the situation is more complex in the presence of jumps. In conclusion, the findings in this paper indicate that a way to obtain a larger number of rate optimal and efficient estimators might be to relax the smoothness conditions, and to allow the estimating function to discriminate more explicitly between intervals with jumps and intervals without jumps.

Let us illustrate this conclusion by a brief consideration of the maximum likelihood estimator. In the particular case where  $c(x, z, \alpha) = z$ , the approximation to the transition density function proposed in [63] can be applied. If we assume that the drift function depends on  $\alpha_1$  and the jump mechanism on  $\alpha_2$ , we find the approximation

$$p(\Delta, y, x, \theta) = \frac{L(y, x, \alpha_1, \beta)}{\sqrt{2\pi\Delta}b(y, \beta)}(1 + O(\Delta)) + \varphi(y - x, \alpha_2) + O(\Delta^2),$$

where

$$\begin{aligned} & L(y, x, \alpha_1, \beta) \\ &= \exp \left( -\frac{(f_1(y, \beta) - f_1(x, \beta))^2}{2\Delta} + f_2(y, \alpha_1, \beta) - f_2(x, \alpha_1, \beta) - \frac{1}{2} \log \frac{b(y, \beta)}{b(x, \beta)} \right) \end{aligned}$$

with  $f_1(x, \beta) = \int^x 1/b(z, \beta) dz$  and  $f_2(x, \alpha_1, \beta) = \int^x a(z, \alpha_1)/b^2(z, \beta) dz$ . Note that in the case considered here,  $\varphi(x, w, \alpha_2)$  does not depend on the first argument. Furthermore,  $\Delta^{-1/2} \exp(-(f_1(y, \beta) - f_1(x, \beta))^2/(2\Delta))$  functions as a kind of smooth indicator function which ensures that when increments are large, weight is mainly put on  $\varphi$  when  $\Delta$  is small.

We consider the normalized score function of the form (1.2) with

$$g(\Delta, y, x, \theta) = (\partial_{\alpha_1} \log p(\Delta, y, x, \theta), \partial_{\alpha_2} \log p(\Delta, y, x, \theta), \Delta \partial_\beta \log p(\Delta, y, x, \theta))^*.$$

Under weak regularity conditions, the score function is a martingale, see e.g. [2]. By straightforward, but rather long, calculations it follows from the approximation proposed in [63] that the derivatives of  $g(\Delta, y, x, \theta)$  satisfy the equalities in Condition 4.4 for all  $x$  and  $y$  and the efficiency conditions in Condition 4.6 – under suitable regularity assumptions on the remainder terms. However, the score function does not satisfy the smoothness condition in Assumption 2.6.(ii). Specifically, the function  $y \mapsto g(0, y, x, \theta)$  is not continuous at  $y = x$ . Therefore, for instance,  $\partial_y g_\alpha(0, x, x, \theta)$  in Condition 4.6 must be interpreted as  $\lim_{\Delta \rightarrow 0} \partial_y g_\alpha(\Delta, x, x, \theta)$ . This discontinuity explains why  $g$  can satisfy both Condition 4.4 and Condition 4.6. More importantly, the proofs of the present paper do not apply to the maximum likelihood

estimator. Whereas our proofs and smoothness conditions are natural for the usual approximate martingale estimating functions based on conditional moments, it is an interesting theoretical question whether the proofs can be modified to cover less smooth estimating functions like the score function.

## 5 Proofs

Section 5.1 presents lemmas used in Section 5.2, together with results from [22], to prove Theorems 3.2 and 4.5. Proofs of the lemmas are given in Section 5.3.

### 5.1 Main Lemmas

**Lemma 5.1.** *Consider the model given by (1.1). Suppose that Assumptions 2.2 and 2.6 hold. If (2.3) holds with  $R_\theta(t, x)$  not identically equal to 0, suppose, additionally, that  $n\Delta_n^{2\kappa_0-1} \rightarrow 0$  as  $n \rightarrow \infty$ . For  $\theta \in \Theta$ , let*

$$\begin{aligned} A(\theta, \theta_0) &= \int_{\mathcal{X}} (\mathcal{L}(g(0, \theta))(x, x) - \mathcal{L}_\theta(g(0, \theta))(x, x)) \pi(dx) \\ B(\theta, \theta_0) &= \int_{\mathcal{X}} (\mathcal{L}(\partial_\theta g(0, \theta))(x, x) - \partial_\theta \mathcal{L}_\theta(g(0, \theta))(x, x)) \pi(dx) \\ C(\theta, \theta_0) &= \int_{\mathcal{X}} \mathcal{L}(gg^*(0, \theta))(x, x) \pi(dx). \end{aligned}$$

Then,

(i) *the mappings  $\theta \mapsto A(\theta, \theta_0)$ ,  $\theta \mapsto B(\theta, \theta_0)$ , and  $\theta \mapsto C(\theta, \theta_0)$  are continuous on  $\Theta$ , with  $A(\theta_0, \theta_0) = 0$  and  $\partial_\theta A(\theta, \theta_0) = B(\theta, \theta_0)$ .*

(ii) *for all compact, convex sets  $K \subseteq \Theta$ ,*

$$\begin{aligned} \sup_{\theta \in K} \left\| \frac{1}{n\Delta_n} \sum_{i=1}^n g(\Delta_n, X_i^n, X_{i-1}^n, \theta) - A(\theta, \theta_0) \right\| &\xrightarrow{\mathcal{P}} 0, \\ \sup_{\theta \in K} \left\| \frac{1}{n\Delta_n} \sum_{i=1}^n \partial_\theta g(\Delta_n, X_i^n, X_{i-1}^n, \theta) - B(\theta, \theta_0) \right\| &\xrightarrow{\mathcal{P}} 0, \\ \sup_{\theta \in K} \left\| \frac{1}{n\Delta_n} \sum_{i=1}^n gg^*(\Delta_n, X_i^n, X_{i-1}^n, \theta) - C(\theta, \theta_0) \right\| &\xrightarrow{\mathcal{P}} 0. \end{aligned} \tag{5.1}$$

(iii) *it holds that*

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, C(\theta_0, \theta_0)).$$

◇

*Remark 5.2.* Consider the model given by (4.1), and let  $B(\theta_0, \theta_0)$  and  $C(\theta_0, \theta_0)$  be as defined in Lemma 5.1. Under Assumptions 2.2, 2.6, and 4.1, we may write

$$\begin{aligned} B(\theta_0, \theta_0) = & - \int_{\mathcal{X}} (\partial_y g(0, x, x, \theta_0) \partial_\theta a(x, \alpha_0) + \frac{1}{2} \partial_y^2 g(0, x, x, \theta_0) \partial_\theta b^2(x, \beta_0)) \pi(dx) \\ & - \int_{\mathcal{X}} \int_{\mathcal{W}(x)} g(0, x+w, x, \theta_0) \partial_\theta \varphi(x, w, \alpha_0) \eta_x(dw) \pi(dx), \end{aligned}$$

and

$$\begin{aligned} C(\theta_0, \theta_0) = & \int_{\mathcal{X}} b^2(x, \beta_0) \partial_y g(\partial_y g)^*(0, x, x, \theta_0) \pi(dx) \\ & + \int_{\mathcal{X}} \int_{\mathcal{W}(x)} g g^*(0, x+w, x, \theta_0) \varphi(x, w, \alpha_0) \eta_x(dw) \pi(dx). \end{aligned}$$

◻

**Lemma 5.3.** Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumptions 2.2, 2.6, and Condition 4.4 hold. If (2.3) holds with  $R_\theta(t, x)$  not identically equal to 0, we suppose, moreover, that  $n \Delta_n^{2(\kappa_0-1)} \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $\delta_n = \text{diag}(\sqrt{n \Delta_n}, \sqrt{n \Delta_n}, \sqrt{n})$ , and define the block diagonal matrix  $D(\theta_0, \theta_0) = \text{diag}(D_1(\theta_0, \theta_0), D_2(\theta_0, \theta_0))$ , where  $D_1$  and  $D_2$  are defined in Theorem 4.5. Then,

(i) for all compact, convex sets  $K \subseteq \Theta$ ,

$$\begin{aligned} \sup_{\theta \in K} \left\| \frac{1}{n \Delta_n^{3/2}} \sum_{i=1}^n \partial_\alpha g_\beta(\Delta_n, X_i^n, X_{i-1}^n, \theta) \right\| & \xrightarrow{\mathcal{P}} 0, \\ \sup_{\theta \in K} \left| \frac{1}{n \Delta_n^2} \sum_{i=1}^n g_\beta^2(\Delta_n, X_i^n, X_{i-1}^n, \theta) - D_2(\theta, \theta_0) \right| & \xrightarrow{\mathcal{P}} 0. \end{aligned} \quad (5.2)$$

(ii) it holds that

$$\delta_n \frac{1}{n \Delta_n} \sum_{i=1}^n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, D(\theta_0, \theta_0)).$$

◊

## 5.2 Proofs of Main Theorems

**Proof of Theorem 3.2.** By Assumption 3.1 and Lemma 5.1,  $G_n(\theta)$  satisfies [22, Condition 2.2] with  $G(\theta) = A(\theta, \theta_0)$  for all compact, convex sets  $K \subseteq \Theta$  with  $\theta_0 \in \text{int } K$ , as well as [22, Condition 2.10] with  $A_n = \sqrt{n \Delta_n} I_d$ ,  $Z = \mathcal{N}_d(0, C(\theta_0, \theta_0))$ , and  $W(\theta) = B(\theta, \theta_0)$ .

By [22, Theorem 2.5], there exists a consistent  $G_n$ -estimator  $\hat{\theta}_n$ . If  $\tilde{\theta}_n$  is any  $G_n$ -estimator which satisfies that  $\mathbb{P}(\tilde{\theta}_n \in K) \rightarrow 1$  as  $n \rightarrow \infty$ , then  $\tilde{\theta}_n$  is consistent by [22, Theorem 2.7.(1)], and Theorem 3.2.(i) is a consequence of the uniqueness result in [22, Theorem 2.5]. Theorem 3.2.(ii) follows from [22, Theorem 2.11], while Theorem 3.2.(iii) follows by properties of convergence in probability, and the continuous mapping theorem. ◻

**Proof of Theorem 4.5.** Let  $B(\theta, \theta_0)$  and  $D(\theta_0, \theta_0)$  be as given in Lemmas 5.1 and 5.3, respectively, and define the block diagonal matrix

$$B_0(\theta, \theta_0) = \text{diag}((B_{jk}(\theta, \theta_0))_{j,k \in \{1,2\}}, B_{33}(\theta, \theta_0)).$$

Note that  $B_0(\theta_0, \theta_0)$  equals  $\text{diag}(B_1(\theta_0, \theta_0), B_2(\theta_0, \theta_0))$ , and that  $D_1(\theta_0, \theta_0)$  is equal to  $(C_{jk}(\theta_0, \theta_0))_{j,k \in \{1,2\}}$  of Lemma 5.1.

As noted in the proof of Theorem 3.2,  $G_n(\theta)$  satisfies [22, Condition 2.2] with  $G(\theta) = A(\theta, \theta_0)$  for all compact, convex sets  $K \subseteq \Theta$  with  $\theta_0 \in \text{int } K$ . Additionally, by Lemmas 5.1 and 5.3, [22, Condition 2.10] is satisfied with  $A_n = \text{diag}(\sqrt{n\Delta_n}, \sqrt{n\Delta_n}, \sqrt{n})$ ,  $Z = \mathcal{N}_3(0, D(\theta_0, \theta_0))$ , and  $W(\theta) = B_0(\theta, \theta_0)$ , so (4.7) follows from [22, Theorem 2.11]. The proof is completed by application of properties of convergence in probability and the continuous mapping theorem.  $\square$

### 5.3 Proofs of Main Lemmas

In the following, we use the notation  $\mathbb{E}_{i-1}^n(\cdot) = \mathbb{E}(\cdot \mid X_{i-1}^n)$ . A martingale difference central limit theorem [18, Corollary 3.1] is used several times without reference.

**Proof of Lemma 5.1.** The smoothness and polynomial growth assumptions on the integrands of  $A(\theta, \theta_0)$ ,  $B(\theta, \theta_0)$ , and  $C(\theta, \theta_0)$ , as well as Assumption 2.2.(iv), ensure that the results in Lemma 5.1.(i) hold.

To prove Lemma 5.1.(ii), use Lemmas A.10 and 2.3 to see that for  $j, k = 1, \dots, d$ ,

$$\begin{aligned} & \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_j(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= \frac{1}{n} \sum_{i=1}^n (\mathcal{L}(g_j(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \mathcal{L}_\theta(g_j(0, \theta))(X_{i-1}^n, X_{i-1}^n)) \\ & \quad + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \\ & \xrightarrow{\mathcal{P}} A_j(\theta, \theta_0), \\ & \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n (\partial_{\theta_k} g_j(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= \frac{1}{n} \sum_{i=1}^n (\mathcal{L}(\partial_{\theta_k} g_j(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \partial_{\theta_k} \mathcal{L}_\theta(g_j(0, \theta))(X_{i-1}^n, X_{i-1}^n)) \\ & \quad + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \\ & \xrightarrow{\mathcal{P}} B_{jk}(\theta, \theta_0), \end{aligned}$$

$$\begin{aligned}
& \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_j g_k(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \frac{1}{n} \sum_{i=1}^n \mathcal{L}(g_j g_k(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \\
&\xrightarrow{\mathcal{P}} C_{jk}(\theta, \theta_0),
\end{aligned} \tag{5.3}$$

and

$$\begin{aligned}
& \frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_j^2(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0, \\
& \frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \mathbb{E}_{i-1}^n ((\partial_{\theta_k} g_j)^2(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0, \\
& \frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_j^2 g_k^2(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0.
\end{aligned}$$

Consequently, by [10, Lemma 9], it holds that pointwise for  $\theta \in \Theta$ ,

$$\begin{aligned}
A_j^n(\theta) &:= \frac{1}{n\Delta_n} \sum_{i=1}^n g_j(\Delta_n, X_i^n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} A_j(\theta, \theta_0), \\
B_{jk}^n(\theta) &:= \frac{1}{n\Delta_n} \sum_{i=1}^n \partial_{\theta_k} g_j(\Delta_n, X_i^n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} B_{jk}(\theta, \theta_0), \\
C_{jk}^n(\theta) &:= \frac{1}{n\Delta_n} \sum_{i=1}^n g_j g_k(\Delta_n, X_i^n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} C_{jk}(\theta, \theta_0).
\end{aligned}$$

Let a compact, convex set  $K \subseteq \Theta$  be given. The functions  $g_j$ ,  $\partial_{\theta_k} g_j$ , and  $g_j g_k$  satisfy the conditions of Lemma A.6, which may be used, together with Jensen's inequality, the mean value theorem, and the Cauchy-Schwarz inequality, to conclude the existence of constants  $p > d$  and  $C_{K,p} > 0$  such that for all  $\theta, \theta' \in K$ ,

$$\mathbb{E} (|H^n(\theta) - H(\theta, \theta_0) - H^n(\theta') + H(\theta', \theta_0)|^p) \leq C_{K,p} \|\theta - \theta'\|^p$$

for  $(H^n, H)$  equal to  $(A_j^n, A_j)$ ,  $(B_{jk}^n, B_{jk})$ , and  $(C_{jk}^n, C_{jk})$ ,  $j, k \in \{1, 2, \dots, d\}$ . Now, using results on convergence in probability from [26, Chapter 14], it holds that

$$\sup_{\theta \in K} |H^n(\theta) - H(\theta, \theta_0)| \xrightarrow{\mathcal{P}} 0$$

as  $n \rightarrow \infty$ , and the results in Lemma 5.1.(ii) follow.

In order to prove Lemma 5.1.(iii), suppose first that the estimating function is a martingale estimating function, i.e. that  $R_\theta(t, x) \equiv 0$  in (2.3). Let  $v \in \mathbb{R}^d$  be a fixed vector, and consider

$$M_{n,i} = \frac{1}{\sqrt{n\Delta_n}} \sum_{j=1}^i v^\star g(\Delta_n, X_j^n, X_{j-1}^n, \theta_0), \tag{5.4}$$



which constitutes a real-valued, zero-mean, square-integrable martingale array with martingale differences  $D_{n,i} = (n\Delta_n)^{-1/2}v^*g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)$ . By (5.3), it holds that

$$\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n ((v^*g)^2(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \xrightarrow{\mathcal{P}} v^*C(\theta_0, \theta_0)v.$$

Furthermore, the conditional Lyapunov condition

$$\frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \mathbb{E}_{i-1}^n ((v^*g)^4(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \xrightarrow{\mathcal{P}} 0 \quad (5.5)$$

holds, implying the Lindeberg condition. The convergence in (5.5) may be verified by multiplying out the parentheses in the left-hand side of the expression and using Lemmas A.10 and 2.3. It follows that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n v^*g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^*C(\theta_0, \theta_0)v),$$

and, by the Cramér-Wold device, that Lemma 5.1.(iii) holds for martingale estimating functions.

If the estimating function is not a martingale estimating function, i.e. if (2.3) holds with  $R_\theta(\Delta_n, X_{i-1}^n)$  not identically 0, it is assumed that  $n\Delta_n^{2\kappa_0-1} \rightarrow 0$  as  $n \rightarrow \infty$  for some  $\kappa_0 \geq 2$ . Let

$$\tilde{g}(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) = g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) - \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)). \quad (5.6)$$

Since

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) = \sqrt{n\Delta_n^{\kappa_0-1/2}} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n) \xrightarrow{\mathcal{P}} 0$$

by (2.3) and Lemma 2.3, it remains to show that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \tilde{g}(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, C(\theta_0, \theta_0)). \quad (5.7)$$

Consider (5.4) with  $g$  replaced by  $\tilde{g}$ . By (2.3) and (5.3),

$$\begin{aligned} & \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n ((v^*\tilde{g})^2(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \\ &= v^* \left( \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g\tilde{g}^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \right) v \\ & \quad - v^* \left( \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0))^* \right) v \end{aligned}$$

$$\xrightarrow{\mathcal{P}} v^* C(\theta_0, \theta_0) v.$$

Furthermore, the conditional Lyapunov condition (5.5) with  $g$  replaced by  $\tilde{g}$  holds as well, and may be verified in the same manner as (5.5). Thus, (5.7) follows, proving Lemma 5.1.(iii) also for approximate martingale estimating functions.  $\square$

**Proof of Lemma 5.3.** Let

$$D_2(\theta, \theta_0) = \int_{\mathcal{X}} \frac{1}{2} \left( b^4(x, \beta_0) + \frac{1}{2} (b^2(x, \beta_0) - b^2(x, \beta))^2 \right) (\partial_y^2 g_3(0, x, x, \theta))^2 \pi(dx).$$

First, use Lemmas A.13, A.14, and 2.3 to see that

$$\begin{aligned} & \frac{1}{n\Delta_n^2} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_3^2(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2} b^4(X_{i-1}^n, \beta_0) (\partial_y^2 g_3(0, X_{i-1}^n, X_{i-1}^n, \theta))^2 \\ & \quad + \frac{1}{n} \sum_{i=1}^n \frac{1}{4} (b^2(X_{i-1}^n, \beta_0) - b^2(X_{i-1}^n, \beta))^2 (\partial_y^2 g_3(0, X_{i-1}^n, X_{i-1}^n, \theta))^2 \\ & \quad + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \\ & \xrightarrow{\mathcal{P}} D_2(\theta, \theta_0), \end{aligned} \tag{5.8}$$

$$\frac{1}{n^2\Delta_n^4} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_3^4(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0,$$

and that for  $j = 1, 2$

$$\frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n \mathbb{E}_{i-1}^n (g_j g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \Delta_n^{1/2} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_i^n, \theta) \xrightarrow{\mathcal{P}} 0, \tag{5.9}$$

$$\begin{aligned} & \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n \mathbb{E}_{i-1}^n (\partial_{\alpha_j} g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \Delta_n^{1/2} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0, \\ & \frac{1}{n^2\Delta_n^3} \sum_{i=1}^n \mathbb{E}_{i-1}^n \left( (\partial_{\alpha_j} g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta))^2 \right) = \frac{1}{n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0. \end{aligned}$$

Thus,

$$\begin{aligned} & \frac{1}{n\Delta_n^2} \sum_{i=1}^n g_3^2(\Delta_n, X_i^n, X_{i-1}^n, \theta) - D_2(\theta, \theta_0) \xrightarrow{\mathcal{P}} 0, \\ & \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n \partial_{\alpha_j} g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta) \xrightarrow{\mathcal{P}} 0 \end{aligned} \tag{5.10}$$

pointwise for  $\theta \in \Theta$ , by [10, Lemma 9]. The functions  $\partial_{\alpha_j} g_3$  and  $g_3^2$ , respectively, satisfy the conditions on  $f$  in Lemmas A.7 and A.8. Consequently, Lemma 5.3.(i) follows from (5.10), Jensen's inequality, the mean value theorem, the Cauchy-Schwarz inequality, as well as results on convergence in probability from [26, Chapter 14].

In order to prove Lemma 5.3.(ii), first observe that

$$\begin{aligned} & \frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \delta_n \mathbb{E}_{i-1}^n (gg^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \delta_n \\ &= \mathbb{E}_{i-1}^n \left( \begin{pmatrix} \frac{1}{n\Delta_n} \sum_{i=1}^n g_\alpha g_\alpha^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) & \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n g_\alpha g_\beta(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \\ \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n g_\beta g_\alpha^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) & \frac{1}{n\Delta_n^2} \sum_{i=1}^n g_\beta^2(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \end{pmatrix} \right), \end{aligned}$$

so combining (5.3) and Remark A.9 for the submatrix concerning  $g_\alpha g_\alpha^*$ , and (5.8) and (5.9) for the remaining coordinates, it follows that

$$\frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \delta_n \mathbb{E}_{i-1}^n (gg^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \delta_n \xrightarrow{\mathcal{P}} D(\theta_0, \theta_0). \quad (5.11)$$

Suppose now that  $R_\theta(t, x) \equiv 0$  in (2.3). Let  $v \in \mathbb{R}^3$  be fixed, and consider

$$M_{n,i} = \frac{1}{n\Delta_n} \sum_{j=1}^i v^* \delta_n g(\Delta_n, X_j^n, X_{j-1}^n, \theta_0) \quad (5.12)$$

which defines a real-valued, zero-mean, square-integrable martingale array with differences  $D_{n,i} = (n\Delta_n)^{-1} v^* \delta_n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)$ . By (5.11), it holds that

$$\sum_{i=1}^n \mathbb{E}_{i-1}^n \left( ((n\Delta_n)^{-1} v^* \delta_n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0))^2 \right) \xrightarrow{\mathcal{P}} v^* D(\theta_0, \theta_0) v.$$

Furthermore, the conditional Lyapunov condition

$$\sum_{i=1}^n \mathbb{E}_{i-1}^n \left( ((n\Delta_n)^{-1} v^* \delta_n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0))^4 \right) \xrightarrow{\mathcal{P}} 0 \quad (5.13)$$

holds, implying also the Lindeberg condition. The Lyapunov condition may be verified by multiplying out the parentheses on the left-hand side of (5.13), and using (A.22), and Lemmas A.14 and 2.3. It follows then that

$$\frac{1}{n\Delta_n} \sum_{i=1}^n v^* \delta_n g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, v^* D(\theta_0, \theta_0) v),$$

thus proving Lemma 5.3.(ii) when  $R_\theta(t, x) \equiv 0$  in (2.3).

When  $R_\theta(t, x)$  is not identically equal to 0 in (2.3), it is assumed that  $n\Delta_n^{2(\kappa_0-1)} \rightarrow 0$  as  $n \rightarrow \infty$  for some  $\kappa_0 \geq 2$ . In this case, define  $\tilde{g}$  as in (5.6). It holds that

$$\frac{1}{n\Delta_n} \sum_{i=1}^n \delta_n \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) = \delta_n \Delta_n^{\kappa_0-1} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n) \xrightarrow{\mathcal{P}} 0,$$

so it remains to show that

$$\frac{1}{n\Delta_n} \sum_{i=1}^n \delta_n \tilde{g}(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, D(\theta_0, \theta_0)). \quad (5.14)$$

Consider (5.12) with  $\tilde{g}$  in place of  $g$ . First, see that

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E}_{i-1}^n \left( \left( (n\Delta_n)^{-1} v^* \delta_n \tilde{g}(\Delta_n, X_i^n, X_{i-1}^n, \theta_0) \right)^2 \right) \\ &= v^* \left( \frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \delta_n \mathbb{E}_{i-1}^n (g g^*(\Delta_n, X_i^n, X_{i-1}^n, \theta_0)) \delta_n \right) v \\ & \quad - v^* \delta_n \Delta_n^{\kappa_0-1} \left( \frac{1}{n^2} \sum_{i=1}^n R(\Delta_n, X_{i-1}^n) \right) \delta_n \Delta_n^{\kappa_0-1} v \\ & \xrightarrow{\mathcal{P}} v^* D(\theta_0, \theta_0) v. \end{aligned}$$

Also, the conditional Lyapunov condition (5.13) holds with  $\tilde{g}$  in place of  $g$ . This is seen by multiplying out the parentheses in the conditional expectation, and using Lemmas A.10, A.13, A.14, and Lemma 2.3. Now (5.14) follows, completing the proof of Lemma 5.3.(ii).  $\square$

## A Auxiliary Results

This appendix contains technical results pertaining to the proofs in Section 5.3. When not otherwise mentioned, the general model given by (1.1) is assumed. Appendix A.1 contains inequalities involving expectations, most of them used to prove uniform convergence in probability, while Appendix A.2 concerns expansions of conditional moments in terms of the infinitesimal generator (2.2).

### A.1 Inequalities

**Assumption A.1.** *The function  $f(t, y, x, \theta)$ ,  $f : (0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \rightarrow \mathbb{R}$ , and its mixed partial derivatives  $\partial_t^i \partial_y^j f$ ,  $i = 0, 1$ ,  $j = 0, 1, 2$ , exist, are continuous, and are of polynomial growth in  $x$  and  $y$ , uniformly for  $t \in (0, \Delta_0)_{\varepsilon_0}$  and  $\theta$  in compact, convex sets. Furthermore,  $f(0, x, x, \theta) = 0$  for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ .  $\diamond$*

**Definition A.2.** For  $q \in \mathbb{N}$ , let  $\mathbf{z}_q = (z_1, \dots, z_q)^* \in \mathbb{R}^q$  with the convention  $\mathbf{z}_0 = ()$ . Let  $m \in \mathbb{N}_0$  be given. Suppose that Assumption 2.2 holds, and that the function

$(t, y, x, \theta) \mapsto f(t, y, x, \mathbf{z}_m, \theta)$  satisfies Assumption A.1. Define  $\mathcal{A}_1$ ,  $\mathcal{A}_2$ , and  $\mathcal{A}_3$  by their actions on  $f$ , which result in the functions

$$\begin{aligned}\mathcal{A}_1 f &: (t, y, x, \mathbf{z}_m, \theta) \mapsto \partial_t f(t, y, x, \mathbf{z}_m, \theta) + \mathcal{L}f(t, y, x, \mathbf{z}_m, \theta) \\ \mathcal{A}_2 f &: (t, y, x, \mathbf{z}_m, \theta) \mapsto b(y, \theta_0) \partial_y f(t, y, x, \mathbf{z}_m, \theta) \\ \mathcal{A}_3 f &: (t, y, x, \mathbf{z}_{m+1}, \theta) \mapsto f(t, y + c(y, z_{m+1}, \theta_0), x, \mathbf{z}_m, \theta) - f(t, y, x, \mathbf{z}_m, \theta).\end{aligned}$$

◇

*Remark A.3.* When well-defined for  $f(t, y, x, \mathbf{z}_m, \theta)$ , the notation  $f_j := \mathcal{A}_j f$ ,  $f_{jk} := \mathcal{A}_k \mathcal{A}_j f$ , and  $f_{j3k} := \mathcal{A}_k \mathcal{A}_3 \mathcal{A}_j f$  is used for  $j, k = 1, 2, 3$ . We put  $\mathbf{h}(u, s, \mathbf{z}_m, \theta) := h(u - s, X_u, X_s, \mathbf{z}_m, \theta)$ , as well as  $\mathbf{h}(u-, s, \mathbf{z}_m, \theta) := h(u - s, X_{u-}, X_s, \mathbf{z}_m, \theta)$ , and  $D\mathbf{h}(\cdot, \theta, \theta') := \mathbf{h}(\cdot, \theta) - \mathbf{h}(\cdot, \theta')$  for functions  $h(t, y, x, \mathbf{z}_m, \theta)$ . ○

**Lemma A.4.** *Suppose that Assumption 2.2 holds, and that  $f(t, y, x, \theta)$  satisfies Assumption A.1. Let  $p = 2^q$  for some  $q \in \mathbb{N}$ , and recall the notation from Remark A.3. Then, there exist constants  $C_p > 0$  such that for all  $\theta, \theta' \in \Theta$  and  $n \in \mathbb{N}$ ,*

$$\begin{aligned}& \mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^p \right) \\ & \leq (n\Delta_n)^{p-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_1(u, t_{i-1}^n, \theta, \theta')^p) du \\ & \quad + (n\Delta_n)^{p/2-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_2(u, t_{i-1}^n, \theta, \theta')^p) du \\ & \quad + \sum_{l=1}^q (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du.\end{aligned}$$

◇

**Proof of Lemma A.4.** By Itô's formula for SDEs with jumps,

$$\begin{aligned}& \mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^p \right) \\ & \leq C_p \mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_1(u-, t_{i-1}^n, \theta, \theta') du \right|^p \right) \tag{A.1}\end{aligned}$$

$$+ C_p \mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_2(u-, t_{i-1}^n, \theta, \theta') dW_u \right|^p \right) \tag{A.2}$$

$$+ C_p \mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta') (N^{\theta_0} - \mu_{\theta_0})(du, dz) \right|^p \right) \tag{A.3}$$

for suitable constants  $C_p > 0$ . Starting with (A.1), using Jensen's inequality twice,

$$\mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_1(u-, t_{i-1}^n, \theta, \theta') du \right|^p \right)$$

$$\begin{aligned}
&= (n\Delta_n)^p \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_1(u-, t_{i-1}^n, \theta, \theta') du \right|^p \right) \\
&\leq (n\Delta_n)^{p-1} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_1(u-, t_{i-1}^n, \theta, \theta')^p) du. \tag{A.4}
\end{aligned}$$

Now, consider (A.2). In the same manner as before, using also the Burkholder-Davis-Gundy inequality,

$$\begin{aligned}
&\mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_2(u-, t_{i-1}^n, \theta, \theta') dW_u \right|^p \right) \\
&= (n\Delta_n)^p \mathbb{E} \left( \left| \frac{1}{n} \sum_{i=1}^n \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_2(u-, t_{i-1}^n, \theta, \theta') dW_u \right|^p \right) \\
&\leq (n\Delta_n)^p C_p \mathbb{E} \left( \left| \frac{1}{n^2} \sum_{i=1}^n \frac{1}{\Delta_n^2} \int_{t_{i-1}^n}^{t_i^n} D\mathbf{f}_2(u-, t_{i-1}^n, \theta, \theta')^2 du \right|^{p/2} \right) \\
&\leq (n\Delta_n)^{p/2-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_2(u-, t_{i-1}^n, \theta, \theta')^p) du, \tag{A.5}
\end{aligned}$$

for some constant  $C_p > 0$ . Finally, in the case of (A.3), let  $\mathbf{M}^{(k)} = (M_v^{(k)})_{v \geq 0}$  and  $\mathbf{S}^{(k)} = (S_v^{(k)})_{v \geq 0}$  be given by

$$\begin{aligned}
M_v^{(k)} &= \int_0^v \int_{\mathbb{R}} \sum_{i=1}^n \mathbf{1}_{(t_{i-1}^n, t_i^n]}(u) D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^k (N^{\theta_0} - \mu_{\theta_0})(du, dz) \\
S_v^{(k)} &= \int_0^v \int_{\mathbb{R}} \sum_{i=1}^n \mathbf{1}_{(t_{i-1}^n, t_i^n]}(u) D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^k \nu_{\theta_0}(dz) du \tag{A.6}
\end{aligned}$$

for  $k \in \mathbb{N}$ , and note that the quadratic variation of  $\mathbf{M}^{(k)}$  may be written as

$$\int_0^v \int_{\mathbb{R}} \sum_{i=1}^n \mathbf{1}_{(t_{i-1}^n, t_i^n]}(u) D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^{2k} N^{\theta_0}(du, dz) = M_v^{(2k)} + S_v^{(2k)}.$$

By the Burkholder-Davis-Gundy inequality, for any  $m \geq 1$ , there exist constants  $C_m > 0$  such that

$$\mathbb{E} \left( |M_v^{(k)}|^m \right) \leq C_m \mathbb{E} \left( \left( M_v^{(2k)} \right)^{m/2} \right) + C_m \mathbb{E} \left( \left( S_v^{(2k)} \right)^{m/2} \right).$$

In particular, inserting  $2^j$  in place of  $k$  and  $2^{q-j}$  in place of  $m$  for  $j \in \{0, 1, \dots, q-1\}$ ,

$$\mathbb{E} \left( \left( M_v^{(2^j)} \right)^{2^{q-j}} \right) \leq C_p \mathbb{E} \left( \left( M_v^{(2^{j+1})} \right)^{2^{q-(j+1)}} \right) + C_p \mathbb{E} \left( \left( S_v^{(2^{j+1})} \right)^{2^{q-(j+1)}} \right).$$

This inequality may be used iteratively to obtain

$$\mathbb{E} \left( \left( M_v^{(1)} \right)^p \right) \leq C_p \sum_{l=1}^q \mathbb{E} \left( \left( S_v^{(2^l)} \right)^{2^{q-l}} \right),$$

where we used that  $\mathbb{E}(M_v^{(p)}) = 0$  by properties of the Poisson integral. Inserting from (A.6), this may also be written as

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta') (N^{\theta_0} - \mu_{\theta_0})(du, dz) \right|^p \right) \\ & \leq C_p \sum_{l=1}^q \mathbb{E} \left( \left( \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^{2^l} \nu_{\theta_0}(dz) du \right)^{2^{q-l}} \right). \end{aligned} \quad (\text{A.7})$$

Recalling that  $\nu_{\theta}$  has density  $\xi(\theta)p(\cdot, \theta)$  with respect to  $\tilde{\nu}$ , where  $p(\cdot, \theta)$  is a probability density, Jensen's inequality is used twice to write

$$\begin{aligned} & \mathbb{E} \left( \left( \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^{2^l} \nu_{\theta_0}(dz) du \right)^{2^{q-l}} \right) \\ & = (\xi(\theta_0)n\Delta_n)^{2^{q-l}} \\ & \quad \times \mathbb{E} \left( \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^{2^l} p(z, \theta_0) \tilde{\nu}(dz) du \right)^{2^{q-l}} \right) \\ & \leq (\xi(\theta_0)n\Delta_n)^{2^{q-l}-1} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du \\ & = (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du. \end{aligned}$$

Inserting this into (A.7), we obtain

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta') (N^{\theta_0} - \mu_{\theta_0})(du, dz) \right|^p \right) \\ & \leq \sum_{l=1}^q (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u-, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du. \end{aligned} \quad (\text{A.8})$$

The proof is completed by replacing (A.1), (A.2), (A.3) with (A.4), (A.5), (A.8), and using that  $\mathbf{X}$  has finite activity jumps to replace  $X_{u-}$  with  $X_u$  in the integrals.  $\square$

**Lemma A.5.** *Let  $m \in \mathbb{N}_0$ ,  $p = 2^q$  for some  $q \in \mathbb{N}$ , and recall the notation of Remark A.3. Suppose that Assumption 2.2 holds. Assume that  $(t, y, x, \theta) \mapsto f(t, y, x, \mathbf{z}_m, \theta)$  satisfies Assumption A.1. Then, there exist constants  $C_p > 0$  such that*

$$\begin{aligned} & \mathbb{E} (D\mathbf{f}(t, s, \mathbf{z}_m, \theta, \theta')^p) \\ & \leq (t-s)^{p-1} C_p \int_s^t \mathbb{E} (D\mathbf{f}_1(u, s, \mathbf{z}_m, \theta, \theta')^p) du \\ & \quad + (t-s)^{p/2-1} C_p \int_s^t \mathbb{E} (D\mathbf{f}_2(u, s, \mathbf{z}_m, \theta, \theta')^p) du \\ & \quad + \left( \sum_{l=1}^q (t-s)^{2^{q-l}-1} \right) C_p \int_s^t \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u, s, \mathbf{z}_m, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du \end{aligned}$$

for all  $\theta, \theta' \in \Theta$ ,  $0 \leq s < t \leq s + \Delta_0$ .  $\diamond$

Letting  $f$  depend on an extra variable  $\mathbf{z}_m$  in the proof of Lemma A.4, and putting  $n = 1$ ,  $t_i^n = t$  and  $t_{i-1}^n = s$  (so that  $\Delta_n = t - s$ ) proves Lemma A.5.

**Lemma A.6.** *Let  $p > d$  of the form  $p = 2^q$  for some  $q \in \mathbb{N}$  be given. Suppose that Assumption 2.2 holds, and that  $f(t, y, x, \theta) \in C_{pol}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$  with  $f(0, x, x, \theta) = 0$  for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ . Let*

$$\zeta_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^n f(\Delta_n, X_i^n, X_{i-1}^n, \theta).$$

Then, for each compact, convex set  $K \subseteq \Theta$ , there exists  $C_{K,p} > 0$  such that

$$\mathbb{E}(|\zeta_n(\theta) - \zeta_n(\theta')|^p) \leq C_{K,p} \|\theta - \theta'\|^p$$

for all  $\theta, \theta' \in K$  and  $n \in \mathbb{N}$ .  $\diamond$

**Proof of Lemma A.6.** Recall the notation from Remark A.3. Let  $K \subseteq \Theta$  compact and convex be given. Write

$$\mathbb{E}(|\zeta_n(\theta) - \zeta_n(\theta')|^p) = (n\Delta_n)^{-p} \mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^p \right). \quad (\text{A.9})$$

By Lemma A.4, there exist constants  $C_p > 0$  such that for all  $\theta, \theta' \in K$  and  $n \in \mathbb{N}$ ,

$$\begin{aligned} & \mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^p \right) \\ & \leq (n\Delta_n)^{p-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_1(u, t_{i-1}^n, \theta, \theta')^p) du \\ & \quad + (n\Delta_n)^{p/2-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_2(u, t_{i-1}^n, \theta, \theta')^p) du \\ & \quad + \sum_{l=1}^q (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du. \end{aligned} \quad (\text{A.10})$$

The mean value theorem and the Cauchy-Schwarz inequality may be used to show that there exist constants  $C_{K,p} > 0$  such that for  $j = 1, 2$ ,

$$\int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_j(u, t_{i-1}^n, \theta, \theta')^p) du \leq C_{K,p} \Delta_n \|\theta - \theta'\|^p \quad (\text{A.11})$$

$$\int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z, \theta, \theta')^p) \nu_{\theta_0}(dz) du \leq C_{K,p} \Delta_n \|\theta - \theta'\|^p. \quad (\text{A.12})$$

Inserting (A.11) and (A.12) into (A.10) yields the existence of  $C_{K,p} > 0$  such that

$$\mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^p \right)$$



$$\begin{aligned}
&\leq C_{K,p} \left( (n\Delta_n)^p + (n\Delta_n)^{p/2} + \sum_{l=1}^q (n\Delta_n)^{2^{q-l}} \right) \|\theta - \theta'\|^p \\
&\leq C_{K,p} (n\Delta_n)^p \|\theta - \theta'\|^p,
\end{aligned} \tag{A.13}$$

since  $n\Delta_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Inserting (A.13) into (A.9) completes the proof.  $\square$

**Lemma A.7.** *Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumption 2.2 holds, and that  $f(t, y, x, \theta) \in C_{pol}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ . Furthermore, assume that*

$$\begin{aligned}
f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1, 2 \\
\partial_t f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1 \\
\partial_y f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1 \\
\partial_y^2 f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & \quad k = 0, 1
\end{aligned}$$

for all  $\tilde{\alpha} \in A$ ,  $\theta \in \Theta$ , and  $x \in \mathcal{X}$ , where  $\mathcal{M}_k(y, \alpha)$  is as defined in Section 4.3. Let

$$\zeta_n(\theta) = \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n f(\Delta_n, X_i^n, X_{i-1}^n, \theta).$$

Then, for any compact, convex  $K \subseteq \Theta$ , there exists  $C_K > 0$ , so for all  $\theta, \theta' \in K$ ,  $n \in \mathbb{N}$ ,

$$\mathbb{E} (|\zeta_n(\theta) - \zeta_n(\theta')|^4) \leq C_K \|\theta - \theta'\|^4.$$

$\diamond$

**Proof of Lemma A.7.** Recall the notation of Remark A.3, and note that for  $j = 1, 2, 3$ ,  $f$ ,  $f_j$ , and  $f_{j3}$ , as functions of  $(t, y, x, \theta)$ , satisfy Assumption A.1. Write

$$\mathbb{E} (|\zeta_n(\theta) - \zeta_n(\theta')|^4) = (n\Delta_n)^{-4} \Delta_n^{-2} \mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^4 \right). \tag{A.14}$$

By Lemma A.4, there exist constants  $C > 0$  such that for all  $\theta, \theta' \in \Theta$  and  $n \in \mathbb{N}$ ,

$$\begin{aligned}
&\mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^4 \right) \\
&\leq (n\Delta_n)^3 C \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_1(u, t_{i-1}^n, \theta, \theta')^4) du \\
&\quad + n\Delta_n C \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E} (D\mathbf{f}_2(u, t_{i-1}^n, \theta, \theta')^4) du \\
&\quad + (1 + n\Delta_n) C \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z_1, \theta, \theta')^4) \nu_{\alpha_0}(dz_1) du.
\end{aligned} \tag{A.15}$$

Furthermore, applying Lemma A.5 twice, there exist constants  $C > 0$  such that

$$\begin{aligned}
& \mathbb{E} (D\mathbf{f}_j(u, t_{i-1}^n, \theta, \theta')^4) \\
& \leq C(u - t_{i-1}^n)^3 \int_{t_{i-1}^n}^u \mathbb{E} (D\mathbf{f}_{j1}(v, t_{i-1}^n, \theta, \theta')^4) dv \\
& \quad + C(u - t_{i-1}^n) \int_{t_{i-1}^n}^u \mathbb{E} (D\mathbf{f}_{j2}(v, t_{i-1}^n, \theta, \theta')^4) dv \\
& \quad + C(1 + u - t_{i-1}^n) \\
& \quad \times \left( \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (v - t_{i-1}^n)^3 \int_{t_{i-1}^n}^v \mathbb{E} (D\mathbf{f}_{j31}(w, t_{i-1}^n, z_1, \theta, \theta')^4) dw \nu_{\alpha_0}(dz_1) dv \right. \\
& \quad + \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (v - t_{i-1}^n) \int_{t_{i-1}^n}^v \mathbb{E} (D\mathbf{f}_{j32}(w, t_{i-1}^n, z_1, \theta, \theta')^4) dw \nu_{\alpha_0}(dz_1) dv \\
& \quad + \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (1 + v - t_{i-1}^n) \\
& \quad \times \left. \int_{t_{i-1}^n}^v \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_{j33}(w, t_{i-1}^n, \mathbf{z}_2, \theta, \theta')^4) \nu_{\alpha_0}(dz_2) dw \nu_{\alpha_0}(dz_1) dv \right) \tag{A.16}
\end{aligned}$$

for  $j = 1, 2$ , and

$$\begin{aligned}
& \mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z_1, \theta, \theta')^4) \\
& \leq C(u - t_{i-1}^n)^3 \int_{t_{i-1}^n}^u \mathbb{E} (D\mathbf{f}_{31}(v, t_{i-1}^n, z_1, \theta, \theta')^4) dv \\
& \quad + C(u - t_{i-1}^n) \int_{t_{i-1}^n}^u \mathbb{E} (D\mathbf{f}_{32}(v, t_{i-1}^n, z_1, \theta, \theta')^4) dv \\
& \quad + C(1 + u - t_{i-1}^n) \\
& \quad \times \left( \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (v - t_{i-1}^n)^3 \int_{t_{i-1}^n}^v \mathbb{E} (D\mathbf{f}_{331}(w, t_{i-1}^n, \mathbf{z}_2, \theta, \theta')^4) dw \nu_{\alpha_0}(dz_2) dv \right. \\
& \quad + \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (v - t_{i-1}^n) \int_{t_{i-1}^n}^v \mathbb{E} (D\mathbf{f}_{332}(w, t_{i-1}^n, \mathbf{z}_2, \theta, \theta')^4) dw \nu_{\alpha_0}(dz_2) dv \\
& \quad + \int_{t_{i-1}^n}^u \int_{\mathbb{R}} (1 + v - t_{i-1}^n) \\
& \quad \times \left. \int_{t_{i-1}^n}^v \int_{\mathbb{R}} \mathbb{E} (D\mathbf{f}_{333}(w, t_{i-1}^n, \mathbf{z}_3, \theta, \theta')^4) \nu_{\alpha_0}(dz_3) dw \nu_{\alpha_0}(dz_2) dv \right) . \tag{A.17}
\end{aligned}$$

Let a compact and convex subset  $K \subseteq \Theta$  be given. Using the mean value theorem and the Cauchy-Schwarz inequality, it may be shown that there exist constants  $C_K > 0$  such that for  $i = 1, \dots, n$ , and  $j \in \{11, 12, 21, 22\}$ ,  $k \in \{31, 32, 131, 132, 231, 232\}$ ,

and  $l \in \{133, 233, 331, 332\}$ ,

$$\begin{aligned}
\mathbb{E} (D\mathbf{f}_j(w, t_{i-1}^n, \theta, \theta')^4) &\leq C_K \|\theta - \theta'\|^4 \\
\mathbb{E} (D\mathbf{f}_k(w, t_{i-1}^n, z_1, \theta, \theta')^4) &\leq C_K \|\theta - \theta'\|^4 (1 + |z_1|^{C_K}) \\
\mathbb{E} (D\mathbf{f}_l(w, t_{i-1}^n, \mathbf{z}_2, \theta, \theta')^4) &\leq C_K \|\theta - \theta'\|^4 (1 + |z_1|^{C_K}) (1 + |z_2|^{C_K}) \\
\mathbb{E} (D\mathbf{f}_{333}(w, t_{i-1}^n, \mathbf{z}_3, \theta, \theta')^4) &\leq C_K \|\theta - \theta'\|^4 (1 + |z_1|^{C_K}) (1 + |z_2|^{C_K}) (1 + |z_3|^{C_K}).
\end{aligned} \tag{A.18}$$

Inserting (A.18) into (A.16) and (A.17), it follows that for  $j = 1, 2$ ,

$$\begin{aligned}
\mathbb{E} (D\mathbf{f}_j(u, t_{i-1}^n, \theta, \theta')^4) &\leq C_K (u - t_{i-1}^n)^2 \|\theta - \theta'\|^4 \\
\mathbb{E} (D\mathbf{f}_3(u, t_{i-1}^n, z_1, \theta, \theta')^4) &\leq C_K (u - t_{i-1}^n)^2 (1 + |z_1|^{C_K}) \|\theta - \theta'\|^4.
\end{aligned} \tag{A.19}$$

Now, inserting (A.19) into (A.15) yields the existence of  $C_K > 0$  such that

$$\mathbb{E} \left( \left| \sum_{i=1}^n D\mathbf{f}(t_i^n, t_{i-1}^n, \theta, \theta') \right|^4 \right) \leq C_K (n\Delta_n)^4 \Delta_n^2 \|\theta - \theta'\|^4, \tag{A.20}$$

where we used that  $n\Delta_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Inserting (A.20) into (A.14), the desired result is obtained.  $\square$

The proof of the following Lemma A.8 is very similar to the proof of Lemma A.7, but requires more applications of Lemma A.5 in order to achieve appropriate orders of  $\Delta_n$ . We refer to [24, Section 3.A.3] for the details.

**Lemma A.8.** *Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumption 2.2 holds, and that  $f(t, y, x, \theta) \in C_{pol}^\infty((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ . Furthermore, assume that*

$$\begin{aligned}
f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1, 2, 3, 4 \\
\partial_t f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1, 2, 3 \\
\partial_y f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1, 2, 3 \\
\partial_y^2 f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1, 2, 3 \\
\partial_y^3 f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1 \\
\partial_t \partial_y f(0, y, x, \theta) &= 0, & y \in \mathcal{M}_k(x, \tilde{\alpha}), & & k = 0, 1.
\end{aligned}$$

for all  $\tilde{\alpha} \in A$ ,  $\theta \in \Theta$ , and  $x \in \mathcal{X}$ , where  $\mathcal{M}_k(y, \alpha)$  is as defined in Section 4.3. Let

$$\zeta_n(\theta) = \frac{1}{n\Delta_n^2} \sum_{i=1}^n f(\Delta_n, X_i^n, X_{i-1}^n, \theta).$$

Then, for any compact, convex set  $K \subseteq \Theta$ , there exists a constant  $C_K > 0$  such that

$$\mathbb{E} (|\zeta_n(\theta) - \zeta_n(\theta')|^4) \leq C_K \|\theta - \theta'\|^4$$

for all  $\theta, \theta' \in K$ , and  $n \in \mathbb{N}$ .  $\diamond$

## A.2 Expansion of Conditional Moments

*Remark A.9.* Note that under Assumptions 2.2 and 2.6,

$$\begin{aligned}
& \partial_\theta \mathcal{L}_\theta(g(0, \theta))(x, x) \\
&= \mathcal{L}_\theta(\partial_\theta g(0, \theta))(x, x) + \partial_y g(0, x, x, \theta) \partial_\theta a(x, \theta) + \frac{1}{2} \partial_y^2 g(0, x, x, \theta) \partial_\theta b^2(x, \theta) \\
&\quad + \int_{\mathbb{R}} \partial_y g(0, x + c(x, z, \theta), x, \theta) \partial_\theta c(x, z, \theta) \nu_\theta(dz) \\
&\quad + \int_{\mathbb{R}} g(0, x + c(x, z, \theta), x, \theta) \partial_\theta q(z, \theta) \tilde{\nu}(dz) \\
& \mathcal{L}_\lambda(gg^*(0, \theta))(x, x) \\
&= b^2(x, \lambda) \partial_y g(\partial_y g)^*(0, x, x, \theta) + \int_{\mathbb{R}} gg^*(0, x + c(x, z, \lambda), x, \theta) \nu_\lambda(dz)
\end{aligned}$$

for all  $x \in \mathcal{X}$  and  $\lambda, \theta \in \Theta$ , by (2.2) and Lemma 2.7.  $\circ$

Furthermore, note that under Assumption 2.2, it holds that

$$\mathbb{E} \left( R_\lambda(\Delta, X_{t+\Delta}^\lambda, X_t^\lambda, \theta) \mid X_t^\lambda \right) = R_\lambda(\Delta, X_t^\lambda, \theta) \quad (\text{A.21})$$

for  $0 \leq t \leq t + \Delta$  with  $\Delta \leq \Delta_0$ , and  $\lambda \in \Theta$ . This follows from a version of [55, Proposition 3.1], which may also be shown to hold in the current framework.

**Lemma A.10.** *Suppose that Assumptions 2.2 and 2.6 hold. Then,*

$$\begin{aligned}
& \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \Delta_n (\mathcal{L}(g(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \mathcal{L}_\theta(g(0, \theta))(X_{i-1}^n, X_{i-1}^n)) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta), \\
& \mathbb{E}_{i-1}^n (\partial_\theta g(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \Delta_n (\mathcal{L}(\partial_\theta g(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \partial_\theta \mathcal{L}_\theta(g(0, \theta))(X_{i-1}^n, X_{i-1}^n)) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta), \\
& \mathbb{E}_{i-1}^n (gg^*(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \Delta_n \mathcal{L}(gg^*(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta),
\end{aligned}$$

and, for  $j, k, l, m = 1, \dots, d$ ,

$$\begin{aligned}
& \mathbb{E}_{i-1}^n \left( (\partial_\theta g)^2(\Delta_n, X_i^n, X_{i-1}^n, \theta) \right) = \Delta_n R(\Delta_n, X_{i-1}^n, \theta) \\
& \mathbb{E}_{i-1}^n (g_j g_k g_l(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \Delta_n R(\Delta_n, X_{i-1}^n, \theta) \\
& \mathbb{E}_{i-1}^n (g_j g_k g_l g_m(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \Delta_n R(\Delta_n, X_{i-1}^n, \theta). \quad (\text{A.22})
\end{aligned}$$

$\diamond$

**Proof of Lemma A.10.** Using (2.4), (A.21), Remark A.9, and Lemmas 2.4 and 2.7 coordinate-wise, write

$$\begin{aligned}
& \mathbb{E}_{i-1}^n (g(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \mathbb{E}_{i-1}^n (g(0, X_i^n, X_{i-1}^n, \theta)) + \Delta_n \mathbb{E}_{i-1}^n (g^{(1)}(X_i^n, X_{i-1}^n, \theta)) \\
&\quad + \Delta_n^2 \mathbb{E}_{i-1}^n (R(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= g(0, X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n \mathcal{L}(g(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
&\quad + \Delta_n \left( g^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n R(\Delta_n, X_{i-1}^n, \theta) \right) + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
&= \Delta_n \left( \mathcal{L}(g(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \mathcal{L}_\theta(g(0, \theta))(X_{i-1}^n, X_{i-1}^n) \right) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta), \\
\\
& \mathbb{E}_{i-1}^n (\partial_\theta g(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \mathbb{E}_{i-1}^n (\partial_\theta g(0, X_i^n, X_{i-1}^n, \theta)) + \Delta_n \mathbb{E}_{i-1}^n (\partial_\theta g^{(1)}(X_i^n, X_{i-1}^n, \theta)) \\
&\quad + \Delta_n^2 \mathbb{E}_{i-1}^n (R(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \partial_\theta g(0, X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n \mathcal{L}(\partial_\theta g(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
&\quad + \Delta_n \left( \partial_\theta g^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n R(\Delta_n, X_{i-1}^n, \theta) \right) \\
&= \Delta_n \left( \mathcal{L}(\partial_\theta g(0, \theta))(X_{i-1}^n, X_{i-1}^n) - \partial_\theta \mathcal{L}_\theta(g(0, \theta))(X_{i-1}^n, X_{i-1}^n) \right) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta), \\
\\
& \mathbb{E}_{i-1}^n (gg^*(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= \mathbb{E}_{i-1}^n (gg^*(0, X_i^n, X_{i-1}^n, \theta)) \\
&\quad + \Delta_n \mathbb{E}_{i-1}^n \left( g^{(1)} g^*(0, X_i^n, X_{i-1}^n, \theta) + g(g^{(1)})^*(X_i^n, X_{i-1}^n, \theta) \right) \\
&\quad + \Delta_n^2 \mathbb{E}_{i-1}^n (R(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\
&= gg^*(0, X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n \mathcal{L}(gg^*(0, \theta))(X_{i-1}^n, X_{i-1}^n) \\
&\quad + \Delta_n \left( g^{(1)} g^*(0, X_{i-1}^n, X_{i-1}^n, \theta) + g(g^{(1)})^*(X_{i-1}^n, X_{i-1}^n, \theta) \right) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
&= \Delta_n \mathcal{L}(gg^*(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta).
\end{aligned}$$

The three remaining equalities follow in a similar, more simple manner.  $\square$

Lemmas A.11 and A.12 provide expressions involving the infinitesimal generator (2.2). These lemmas may be used to prove the subsequent lemmas on the expansion of conditional moments. The generalised Leibnitz formula for the  $m$ th derivative of a product is useful for verifying these formulae. For proofs, see [24, Section 3.A.2].

**Lemma A.11.** *Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumptions 2.2 and 2.6, and Condition 4.4 hold. Then, for  $j = 1, 2, 3$ , the following holds for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ .*

$$\mathcal{L}(g_j g_3(0, \theta))(x, x) = 0$$

and, furthermore,

$$\begin{aligned}
& \mathcal{L}^2(g_j g_3(0, \theta))(x, x) \\
&= \frac{3}{2} b^2(x, \beta_0) (2a(x, \alpha_0) + \partial_y b^2(x, \beta_0)) \partial_y g_j \partial_y^2 g_3(0, x, x, \theta) \\
&\quad + \frac{1}{2} b^4(x, \beta_0) (2\partial_y g_j \partial_y^3 g_3 + 3\partial_y^2 g_j \partial_y^2 g_3)(0, x, x, \theta) \\
&\quad + \int_{\mathbb{R}} \frac{1}{2} \left( b^2(x + c(x, z, \alpha_0), \beta_0) + b^2(x, \beta_0) (1 + \partial_y c(x, z, \alpha_0))^2 \right) \\
&\quad \times g_j \partial_y^2 g_3(0, x + c(x, z, \alpha_0), x, \theta) \nu_{\alpha_0}(dz),
\end{aligned}$$

$$\begin{aligned}
& g_j^{(1)}(x, x, \theta) \\
&= -a(x, \alpha) \partial_y g_j(0, x, x, \theta) - \frac{1}{2} b^2(x, \beta) \partial_y^2 g_j(0, x, x, \theta) \\
&\quad - \int_{\mathbb{R}} g_j(0, x + c(x, z, \alpha), x, \theta) \nu_{\alpha}(dz),
\end{aligned}$$

$$\begin{aligned}
& \mathcal{L}(g_j(0, \theta) g_3^{(1)}(\theta))(x, x) \\
&= -\frac{1}{2} a(x, \alpha_0) b^2(x, \beta) \partial_y g_j \partial_y^2 g_3(0, x, x, \theta) \\
&\quad - \frac{1}{4} b^2(x, \beta) b^2(x, \beta_0) \partial_y^2 g_j \partial_y^2 g_3(0, x, x, \theta) \\
&\quad + b^2(x, \beta_0) \partial_y g_j(0, x, x, \theta) \partial_y g_3^{(1)}(x, x, \theta) \\
&\quad + \int_{\mathbb{R}} g_j(0, x + c(x, z, \alpha_0), x, \theta) g_3^{(1)}(x + c(x, z, \alpha_0), x, \theta) \nu_{\alpha_0}(dz),
\end{aligned}$$

$$\begin{aligned}
& \mathcal{L}(g_j^{(1)}(\theta) g_3(0, \theta))(x, x) \\
&= -\frac{1}{2} a(x, \alpha) b^2(x, \beta_0) \partial_y g_j \partial_y^2 g_3(0, x, x, \theta) \\
&\quad - \frac{1}{4} b^2(x, \beta) b^2(x, \beta_0) \partial_y^2 g_j \partial_y^2 g_3(0, x, x, \theta) \\
&\quad - \frac{1}{2} b^2(x, \beta_0) \left( \int_{\mathbb{R}} g_j(0, x + c(x, z, \alpha), x, \theta) \nu_{\alpha}(dz) \right) \partial_y^2 g_3(0, x, x, \theta).
\end{aligned}$$

◇

**Lemma A.12.** Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumptions 2.2 and 2.6, and Condition 4.4 hold. Then, for  $j, k, l = 1, 2, 3$  and  $m = 1, 2$ , the following nine expressions are equal to 0 for all  $x \in \mathcal{X}$  and  $\theta \in \Theta$ :

$$\begin{array}{ll}
\mathcal{L}(g_j g_k g_l g_3(0, \theta))(x, x) & \mathcal{L}(g_j g_k g_3(0, \theta))(x, x) \\
\mathcal{L}(g_j^{(1)}(\theta) g_k g_3^2(0, \theta))(x, x) & \mathcal{L}(g_j g_k g_3(0, \theta) g_3^{(1)}(\theta))(x, x) \\
\mathcal{L}^2(g_j g_3^3(0, \theta))(x, x) & \mathcal{L}(\partial_{\alpha} g_{\beta}(0, \theta))(x, x) \\
\mathcal{L}(\partial_{\alpha_m} g_{\beta}(0, \theta) \partial_{\alpha_m} g_{\beta}^{(1)}(\theta))(x, x) & \partial_{\alpha} \mathcal{L}(g_{\beta}(0, \theta))(x, x) \\
\mathcal{L}^m((\partial_{\alpha} g_{\beta})^2(0, \theta))(x, x) & 
\end{array}$$

◇

**Lemma A.13.** Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumptions 2.2 and 2.6, and Condition 4.4 hold. Then, for  $j = 1, 2, 3$ , the

following holds for all  $n \in \mathbb{N}$ ,  $i = 1, \dots, n$ , and  $\theta \in \Theta$ .

$$\begin{aligned} & \mathbb{E}_{i-1}^n (g_j g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= \Delta_n^2 \left( \frac{1}{2} \mathcal{L}^2(g_j g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) + g_j^{(1)} g_3^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) \right. \\ & \quad \left. + \mathcal{L}(g_j(0, \theta) g_3^{(1)}(\theta))(X_{i-1}^n, X_{i-1}^n) + \mathcal{L}(g_j^{(1)}(\theta) g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) \right) \\ & \quad + \Delta_n^3 R(\Delta_n, X_{i-1}^n, \theta). \end{aligned}$$

◇

**Proof of Lemma A.13.** Lemmas 2.4, 2.7, and A.11 are used to obtain

$$\begin{aligned} & \mathbb{E}_{i-1}^n (g_j g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= \mathbb{E}_{i-1}^n (g_j g_3(0, X_i^n, X_{i-1}^n, \theta)) + \Delta_n \mathbb{E}_{i-1}^n (g_j(0, X_i^n, X_{i-1}^n, \theta) g_3^{(1)}(X_i^n, X_{i-1}^n, \theta)) \\ & \quad + \Delta_n \mathbb{E}_{i-1}^n (g_j^{(1)}(X_i^n, X_{i-1}^n, \theta) g_3(0, X_i^n, X_{i-1}^n, \theta)) \\ & \quad + \frac{1}{2} \Delta_n^2 \mathbb{E}_{i-1}^n (g_j(0, X_i^n, X_{i-1}^n, \theta) g_3^{(2)}(X_i^n, X_{i-1}^n, \theta)) \\ & \quad + \Delta_n^2 \mathbb{E}_{i-1}^n (g_j^{(1)} g_3^{(1)}(X_i^n, X_{i-1}^n, \theta)) \\ & \quad + \frac{1}{2} \Delta_n^2 \mathbb{E}_{i-1}^n (g_j^{(2)}(X_i^n, X_{i-1}^n, \theta) g_3(0, X_i^n, X_{i-1}^n, \theta)) \\ & \quad + \Delta_n^3 \mathbb{E}_{i-1}^n (R(\Delta_n, X_i^n, X_{i-1}^n, \theta)) \\ &= g_j g_3(0, X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n \mathcal{L}(g_j g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) \\ & \quad + \frac{1}{2} \Delta_n^2 \mathcal{L}^2(g_j g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) \\ & \quad + \Delta_n g_j(0, X_{i-1}^n, X_{i-1}^n, \theta) g_3^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) \\ & \quad + \Delta_n^2 \mathcal{L}(g_j(0, \theta) g_3^{(1)}(\theta))(X_{i-1}^n, X_{i-1}^n) \\ & \quad + \Delta_n g_j^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) g_3(0, X_{i-1}^n, X_{i-1}^n, \theta) \\ & \quad + \Delta_n^2 \mathcal{L}(g_j^{(1)}(\theta) g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) \\ & \quad + \frac{1}{2} \Delta_n^2 g_j(0, X_{i-1}^n, X_{i-1}^n, \theta) g_3^{(2)}(X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n^2 g_j^{(1)} g_3^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) \\ & \quad + \frac{1}{2} \Delta_n^2 g_j^{(2)}(X_{i-1}^n, X_{i-1}^n, \theta) g_3(0, X_{i-1}^n, X_{i-1}^n, \theta) + \Delta_n^3 R(\Delta_n, X_{i-1}^n, \theta) \\ &= \frac{1}{2} \Delta_n^2 \mathcal{L}^2(g_j g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 \mathcal{L}(g_j(0, \theta) g_3^{(1)}(\theta))(X_{i-1}^n, X_{i-1}^n) \\ & \quad + \Delta_n^2 \mathcal{L}(g_j^{(1)}(\theta) g_3(0, \theta))(X_{i-1}^n, X_{i-1}^n) + \Delta_n^2 g_j^{(1)} g_3^{(1)}(X_{i-1}^n, X_{i-1}^n, \theta) \\ & \quad + \Delta_n^3 R(\Delta_n, X_{i-1}^n, \theta). \end{aligned}$$

□

The proof of Lemma A.14 utilises Lemma A.12, and is otherwise similar to the proof of Lemma A.13. For details, see [24, Section 3.A.4].

**Lemma A.14.** Consider the model given by (4.1), with  $A \subseteq \mathbb{R}^2$  and  $B \subseteq \mathbb{R}$ . Suppose that Assumptions 2.2 and 2.6, and Condition 4.4 hold. Then, for  $j, k, l = 1, 2, 3$ , the following holds for all  $n \in \mathbb{N}$ ,  $i = 1, \dots, n$ , and  $\theta \in \Theta$ .

$$\mathbb{E}_{i-1}^n (g_j g_k g_l(\Delta_n, X_i^n, X_{i-1}^n, \theta)) = \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta)$$

$$\begin{aligned}
\mathbb{E}_{i-1}^n (g_j g_k g_l g_3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) &= \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
\mathbb{E}_{i-1}^n (g_j g_3^3(\Delta_n, X_i^n, X_{i-1}^n, \theta)) &= \Delta_n^3 R(\Delta_n, X_{i-1}^n, \theta) \\
\mathbb{E}_{i-1}^n (\partial_\alpha g_\beta(\Delta_n, X_i^n, X_{i-1}^n, \theta)) &= \Delta_n^2 R(\Delta_n, X_{i-1}^n, \theta) \\
\mathbb{E}_{i-1}^n \left( (\partial_\alpha g_\beta(\Delta_n, X_i^n, X_{i-1}^n, \theta))^2 \right) &= \Delta_n^3 R(\Delta_n, X_{i-1}^n, \theta)
\end{aligned}$$

◇

## Acknowledgements

We are grateful to the anonymous reviewers for their helpful and inspiring comments and suggestions which have improved this paper. Nina Munkholt Jakobsen was supported by the Danish Council for Independent Research – Natural Science, through a grant to Susanne Ditlevsen. Michael Sørensen was supported by the Center for Research in Econometric Analysis of Time Series funded by the Danish National Research Foundation. The research is part of the Dynamical Systems Interdisciplinary Network funded by the University of Copenhagen Programme of Excellence.

## References

- [1] Bandi, F. M. and Nguyen, T. H. (2003). On the functional estimation of jump-diffusion models. *Journal of Econometrics*, 116(1-2):293–328.
- [2] Barndorff-Nielsen, O. E. and Sørensen, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes. *International Statistical Review*, 62:133–165.
- [3] Becheri, I. G., Drost, F. C., and Werker, B. J. M. (2016). Asymptotic inference for jump diffusions with state-dependent intensity. *Scandinavian Journal of Statistics*, 43(2):520–542.
- [4] Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39.
- [5] Clément, E. and Gloter, A. (2015). Local asymptotic mixed normality property for discretely observed stochastic differential equations driven by stable Lévy processes. *Stochastic Processes and their Applications*, 125:2316–2352.
- [6] De Jong, F., Drost, F. C., and Werker, B. J. M. (2001). A jump-diffusion model for exchange rates in a target zone. *Statistica Neerlandica*, 55(3):270–300.
- [7] Ditlevsen, S. and Greenwood, P. (2013). The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. *Journal of Mathematical Biology*, 67(2):239–259.
- [8] Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111.



- [9] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557.
- [10] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 29(1):119–151.
- [11] Giesecke, K. and Schwenkler, G. (2018). Simulated likelihood estimators for discretely observed jump-diffusions. Preprint. Available at <http://www.gustavo-schwenkler.com/research>.
- [12] Giraudo, M. T. and Sacerdote, L. (1997). Jump-diffusion processes as models for neuronal activity. *Biosystems*, 40(1-2):75–82.
- [13] Gloter, A., Loukianova, D., and Mai, H. (2018). Jump filtering and efficient drift estimation for Lévy-driven SDEs. *The Annals of Statistics*, 46(4):1445–1480.
- [14] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l’Institut Henri Poincaré (B) Probability and Statistics*, 38(5):711–737.
- [15] Golden, M., García-Portugués, E., Sørensen, M., Mardia, K. V., Hamelryck, T., and Hein, J. (2017). A generative angular model of protein structure evolution. *Molecular Biology and Evolution*, 34(8):2085–2100.
- [16] Gonçalves, F. B., Łatuszyński, K. G., and Roberts, G. O. (2017). Exact Monte Carlo likelihood-based inference for jump-diffusion processes. Preprint. Available at: [arXiv:1707.00332](https://arxiv.org/abs/1707.00332).
- [17] Guay, F. and Schwenkler, G. (2017). Efficient parameter estimation for multivariate jump-diffusions. Paris December 2017 Finance Meeting EUROFIDAI - AFFI. Available at <https://ssrn.com/abstract=2926508>.
- [18] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press.
- [19] Hermann, S., Ickstadt, K., and Müller, C. H. (2018). Bayesian prediction for a jump diffusion process – with application to crack growth in fatigue experiments. *Reliability Engineering & System Safety*, 179:83 – 96.
- [20] Jacobsen, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Delta-optimality. *Scandinavian Journal of Statistics*, 28(1):123–149.
- [21] Jacobsen, M. (2002). Optimality and small Delta-optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668.

- [22] Jacod, J. and Sørensen, M. (2018). A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic Processes*, 21(2):415–434.
- [23] Jahn, P., Berg, R. W., Hounsgaard, J., and Ditlevsen, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, 31(3):563–579.
- [24] Jakobsen, N. M. (2015). *Efficient Estimating Functions for Stochastic Differential Equations*. PhD thesis, Faculty of Science, University of Copenhagen. Available at: <http://web.math.ku.dk/noter/filer/phd15nmj.pdf>.
- [25] Jakobsen, N. M. and Sørensen, M. (2017). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. *Bernoulli*, 23(3):1874–1910.
- [26] Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer series in statistics: Probability and its applications. Springer.
- [27] Kawai, R. (2013). Local asymptotic normality property for Ornstein-Uhlenbeck processes with jumps under discrete sampling. *Journal of Theoretical Probability*, 26(4):932–967.
- [28] Kawai, R. and Masuda, H. (2013). Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling. *ESAIM: Probability and Statistics*, 17:13–32.
- [29] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24:211–229.
- [30] Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5(2):299–314.
- [31] Kohatsu-Higa, A., Nualart, E., and Tran, N. K. (2014). LAN property for a simple Lévy process. *Comptes Rendus Mathématique*, 352(10):859–864.
- [32] Kohatsu-Higa, A., Nualart, E., and Tran, N. K. (2017). LAN property for an ergodic diffusion with jumps. *Statistics*, 51:419–454.
- [33] Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.
- [34] Lee, S. S. and Mykland, P. A. (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies*, 21(6):2535–2563.
- [35] Li, C. and Chen, D. (2016). Estimating jump-diffusions using closed-form likelihood expansions. *Journal of Econometrics*, 195(1):51–70.

- [36] Long, H., Ma, C., and Shimizu, Y. (2017). Least squares estimators for stochastic differential equations driven by small Lévy noises. *Stochastic Processes and their Applications*, 127(5):1475–1495.
- [37] Mai, H. (2014). Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 20(2):919–957.
- [38] Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell’Istituto Italiano degli Attuari*, 64:19–47.
- [39] Mancini, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52.
- [40] Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.
- [41] Mancini, C. and Renò, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics*, 160(1):77–92.
- [42] Masuda, H. (2007). Ergodicity and exponential beta-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, 117(1):35–56.
- [43] Masuda, H. (2008). On stability of diffusions with compound-Poisson jumps. *Bulletin of Informatics and Cybernetics*, 40:61–74.
- [44] Masuda, H. (2011). Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test. *RIMS Kôkyûroku 1752*, pages 113–131.
- [45] Masuda, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *The Annals of Statistics*, 41(3):1593–1641.
- [46] Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144.
- [47] Musila, M. and Lánský, P. (1991). Generalized Stein’s model for anatomically complex neurons. *Biosystems*, 25(3):179–191.
- [48] Ogihara, T. and Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, 14(3):189–229.
- [49] Patel, A. and Kosko, B. (2008). Stochastic resonance in continuous and spiking neuron models with Lévy noise. *Neural Networks, IEEE Transactions on*, 19(12):1993–2008.

- [50] Schmisser, É. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications*, 124(1):883–914.
- [51] Shimizu, Y. (2006a). Density estimation of Lévy measures for discretely observed diffusion processes with jumps. *Journal of The Japan Statistical Society*, 36(1):37–62.
- [52] Shimizu, Y. (2006b). M-estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Statistical Inference for Stochastic Processes*, 9(2):179–225.
- [53] Shimizu, Y. (2008). Some remarks on estimation of diffusion coefficients for jump-diffusions from finite samples. *Bulletin of Informatics and Cybernetics*, 40:51–60.
- [54] Shimizu, Y. (2009). Functional estimation for Lévy measures of semimartingales with Poissonian jumps. *Journal of Multivariate Analysis*, 100(6):1073–1092.
- [55] Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, 9:227–277.
- [56] Sørensen, M. (1991). Likelihood methods for diffusions with jumps. In Prabhu, N. U. and Basawa, I. V., editors, *Statistical Inference in Stochastic Processes*, pages 67–105. Marcel Dekker, New York.
- [57] Sørensen, M. (2012). Estimating functions for diffusion-type processes. In Kessler, M., Lindner, A., and Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*, pages 1–107. Chapman & Hall.
- [58] Sørensen, M. (2017). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint. Available at: <http://web.math.ku.dk/~michael/efficient.pdf>.
- [59] Stramer, O., Bognar, M., and Schneider, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. *Journal of Financial Econometrics*, 8(4):450–480.
- [60] Tran, N. K. (2017). LAN property for an ergodic Ornstein–Uhlenbeck process with Poisson jumps. *Communications in Statistics - Theory and Methods*, 46(16):7942–7968.
- [61] Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scandinavian Journal of Statistics*, 31(4):553–566.

- [62] Wang, H. and Zhou, L. (2017). Bandwidth selection of nonparametric threshold estimator in jump–diffusion models. *Computers & Mathematics with Applications*, 73(2):211–219.
- [63] Yu, J. (2007). Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan. *Journal of Econometrics*, 141(2):1245–1280.
- [64] Zhou, L. (2017). Double-smoothed drift estimation of jump-diffusion model. *Communications in Statistics - Theory and Methods*, 46(8):4137–4149.